This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

REMARKS

Reconsideration of this application is respectfully requested.

Claims 35, 37, and 39 have been amended. Claims 47-50 are new and are derived from claims 35-46. No new matter enters by amendment.

Claims 35, 37, 39, 41, 43, and 45 were rejected under 35 U.S.C. § 102(e) as allegedly being anticipated by Chang (U.S. Patent No. 6,001,977) and claims 35-46 were rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Chang et al. in view of White et al. (U.S. Patent No. 4,677,054). The Examiner's position rests on the allegation that applicants' LAV strain and Chang's HTLV-III strain came from the same person and represent different isolates of the same virus. Based on this allegation, the Examiner concludes that differences between Chang's sequence and applicants' sequence are due to "sequencing errors."

Applicants traverse the rejection. There is no evidence of record to support the Examiner's conclusion. Rather, the evidence of record supports the opposite conclusion. First, the sequences of applicants' LAV clone and Chang's HTLV-III clones are different. (*See* U.S. Patent No. 6,001,977.) Second, the extent of differences is higher than would be expected for sequencing errors. (*See* Applicants' August 1, 2003, Amendment at 3-5.) Third, Ratner et al. indicates that differences in the sequence of clones of HIV-1 need not be due to sequencing errors. (Ratner et al. at 59.)

The Examiner's conclusion that Ratner et al. does not support applicants' position is in error. Ratner et al. found a substantial number (79) of differences between two different clones of HTLV-III (HXB2 and BH10). Ratner et al. indicates that these differences are unlikely "to represent cloning artifacts or sequencing errors since 1)

these alterations were confirmed by DNA sequences from both strands of both clones, and 2) 82% of these changes are present in other previously sequenced HTLVIII/LAV clones." (Ratner et al. at 59, ¶ 4.) Since Ratner et al. shows that two different clones of the same strain (HTLV-III) can have substantial differences in sequence, there is no reason to believe that the differences between applicants' LAV clone and Chang's HTLV-III clone must be due to sequencing errors.

Similar to the sequence of BH10 and BH5 in the '977 patent (*see* Table 3 in Applicants' September 3, 2002, Amendment and Response), Ratner's sequence contains a frameshift in the Vpr orf due to an additional "t" at position ~5350. (Ratner et al. at 61, Figure 1.) Thus, three different clones of HTLV-III contain the same difference from applicants' strain. Consequently, Ratner et al. supports that Chang's BH10 clone does not encode applicants' Vpr. Accordingly, applicants' claimed nucleic acids expressing Vpr cannot be anticipated by Chang.

Furthermore, Muesing et al. (Exhibit 1) independently isolated cDNA and proviral clones of HTLV-III from the H9/HTLVIII cell line. (Muesing et al. at 450, col. 2.) The sequence of HTLV-III presented in Figure 3 shows that Muesing's clones, similar to Chang's clones, contain a frameshift in the Vpr orf due to an additional "t" at position ~5350. (*Id.* at 453, Figure 3.) Thus, Muesing et al. provides evidence that the virus or viruses in the H9/HTLVIII cell line, from which Chang's clones were derived, do not encode applicants' Vpr. Accordingly, applicants' claimed nucleic acids expressing Vpr cannot be anticipated by Chang.

Applicants' LAV strain and Chang's HTLV-III strain may have come from the same person and may represent different isolates of the same virus. However, the

issue in this case is whether the sequence of applicants' **clones** and Chang's **clones** are the same. The Examiner's conclusion ignores the potential for the introduction of nucleotide sequence changes during virus replication *in vivo* and *in vitro* during preparation of the clones (*see, e.g.*, Goodenow et al., Peden et al., and Pang et al.; Exhibits 2-4) and the consequent loss of the ability to express particular HIV-1 gene products.

Chang's isolate was passaged in the H9/HTLVIII cell line prior to being cloned. (U.S. Appl. 06/643,306 at 6-7; Exhibit 5.) The H9/HTLVIII cell line was established by infection with material from several patients. (Muesing et al. at 452, col. 1.) Muesing et al. indicates that the H9/HTLVIII cell line contains "about five intact provirus copies." (*Id.*) Moreover, Muesing et al. indicates that "these results suggest that two or more distinct virus isolates are integrated stably in the H9/HTLVIII cell line." (*Id.*) Muesing et al. found 68 nucleotide differences between the sequences of proviral and cDNA clones and concluded that "a significant degree of nucleotide heterogeneity is displayed by the proviral and cDNA sequences." (*Id.*) Consequently, multiple nucleotide sequences of HIV-1 were present in the cell line used in the generation of Chang's clones. In light of this evidence, the differences in sequence between applicants' and Chang's clones are likely due to the presence of multiple viruses in the H9/HTLVIII cell line or due to changes in the viruses during passage in the H9/HTLVIII cell line, and not due to "sequencing errors." Accordingly, applicants respectfully request withdrawal of the rejection.

In addition, generic claim 47 is fully supported in the present application and through the priority chain now claimed to Appln. Ser. No. 06/558,109, filed ***December 5,***

1983. The earliest claimed priority date of Chang is August 22, 1984. Consequently, Chang is not effective prior art with respect to claim 47.

Applicants respectfully submit that this application is in condition for allowance. In the event that the Examiner disagrees, he is invited to call the undersigned to discuss any outstanding issues remaining in this application in order to expedite prosecution.

Please grant any extensions of time required to enter this response and charge any additional required fees to our deposit account 06-0916.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, L.L.P.

Dated: April 23, 2004

By: _____

Salvatore J. Arrigo
Reg. No. 46,063
Telephone: 202-408-4160
Facsimile: 202-408-4400
E-mail: arrigos@finnegan.com

Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus

Mark A. Muesing, Douglas H. Smith, Cirilo D. Cabradilla*, Charles V. Benton†, Laurence A. Lasky‡ & Daniel J. Capon

Departments of Molecular Biology and ‡ Vaccine Development, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

* Center for Infectious Disease, Centers for Disease Control, Atlanta, Georgia 30333, USA

The 9,213-nucleotide structure of the AIDS/lymphadenopathy virus has been determined from molecular clones representing the integrated provirus and viral RNA. The sequence reveals that the virus is highly polymorphic and lacks significant nucleotide homology with type C retroviruses characterized previously. Together with an analysis of the two major viral subgenomic RNAs, these studies establish the coding frames for the gag, pol *and* env *genes and predict the expression of a novel gene at the 3' end of the genome unrelated to the X genes of HTLV-I and -II.*

ACQUIRED immune deficiency syndrome (AIDS) is a novel, transmissible deficiency of cellular immunity characterized by opportunistic infections and certain malignancies, notably *Pneumocystic carinii* pneumonia and Kaposi's sarcoma, in patients without another recognized cause for contracting these rare diseases[1-3]. AIDS is manifested by a profound lymphopenia, a generalized cutaneous anergy and a markedly reduced proliferative response to mitogens, antigens and allogeneic cells, seeming to result from depletion of the OKT4$^+$ T-lymphocyte subset[4]. While humoral immunity is relatively unaffected, there is increasing evidence for a hyperactive B-cell proliferative response which may be related causally to the high incidence of B-lymphoma in AIDS patients[5,6]. In addition to the fully-developed syndrome, an epidemic of a related disease, AIDS-related complex (ARC), has appeared, characterized by generalized chronic lymphadenopathy. This syndrome shares many of the epidemiological features and immune abnormalities and often precedes the clinical manifestations of AIDS. The predominant risk groups for AIDS and ARC include homosexually active males, intravenous drug abusers, recipients of transfusions and blood products and the heterosexual partners and children of high-risk individuals, suggesting the involvement of an infectious agent transmitted through intimate contact or blood products.

Recent evidence has implicated strongly a novel lymphocytotropic retrovirus as the primary aetiological agent of AIDS and the AIDS-related complex. Lymphadenopathy-associated virus (LAV) was isolated initially from cultured lymph-node T cells of patients with lymphadenopathy and AIDS as well as an AIDS patient and an asymptomatic sibling, both with haemophilia B[7-9]. A similar virus, designated human T-lymphotrophic virus type III (HTLV-III), has been isolated from a large number of AIDS and ARC patient blood samples by co-cultivation with the permissive T-cell line H9 (refs 10, 11). LAV and HTLV-III, as well as related retroviruses isolated recently from AIDS patients[12,13], share several important characteristics. Viral replication occurs in the OKT4$^+$ T-lymphocyte population *in vivo* and *in vitro* and is associated with impaired proliferation and the appearance of cytopathic effects[8,10,14]. The virus has a Mg^{2+}-dependent reverse transcriptase, exhibits a dense cylindrical core morphology similar to type D retroviruses[8,13,15] and is recognized by antibodies found in the sera of virtually all AIDS and ARC patients[8,13,16-21].

As a first step towards characterizing the molecular biology of this virus, we have determined the entire nucleotide sequence for one of the integrated proviruses present in H9/HTLV-III cells and for a complete set of overlapping complementary DNAs representing the viral RNA of distinguishable isolate(s) also present in H9/HTLV-III cells. Our results establish that LAV/HTLV-III has no nucleotide homology with previously characterized animal and human retroviruses and that different virus isolates display significant genetic heterogeneity. Together with transcriptional mapping data, these studies provide a detailed picture of the structure and processing of the *gag*, *pol* and *env* gene products, provide evidence for a novel gene in the 3' region of HTLV-III and predict a further gene product in the region between the *pol* and *env* genes.

Isolation of cDNA and provirus clones

Molecular clones of HTLV-III were identified initially from cDNA libraries representing cellular RNA of productively infected H9/HTLV-III cells, established by Popovic *et al.*[10]. The H9 human T-cell line is permissive for the continuous production of high titres of HTLV-III isolated from the cultured lymphocytes of AIDS and ARC patients and is significantly resistant to the cytopathic effects of these viruses. The virus produced by H9/HTLV-III retains its cytopathic activity against fresh normal human lymphocytes[10]. Total poly(A)$^+$RNA was prepared from H9/HTLV-III cells infected with pooled material from several different AIDS patients[10] and used to construct an oligo(dT)-primed cDNA library in the vector λgt10.

The strategy used to identify clones containing HTLV-III sequences was based on differential hybridization with cDNA probes prepared from poly(A)$^+$RNA of H9/HTLV-III cells and the uninfected CEM human T-lymphoblastoid cell line; ~0.2% of the clones in this library contained inserts hybridizing specifically with the H9/HTLV-III cDNA probe. Six of these clones were purified and their DNA inserts used as probes to classify 50 additional H9/HTLV-III-specific clones. Two distinct classes of clones were identified on the basis of their pattern of hybridization; furthermore, weak hybridization was detectable between the two different classes. Inserts from one clone of each H9/HTLV-III-specific class (H9c.7 and H9c.53) were subcloned into phage M13 vectors and their sequences determined by the dideoxy-chain terminator method. Significantly, a 76-nucleotide sequence was shared by the 5' end of H9c.7 and the 3' end of H9c.53, accounting for hybridization between the two classes. The 3' point of divergence of this 76-nucleotide sequence was marked by a polyadenylate tract in H9c.53, as expected for an RNA polymerase II transcript. The congruence exhibited by the opposite ends of these clones was very like the terminal redundancy of the viral genome of retroviruses[22-24], suggesting that clones H9c.7 and H9c.53 represented the 5' and 3' regions, respectively, of HTLV-III (Fig. 1).

The identity of H9c.7 and H9c.53 was confirmed by blot hybridization analysis of H9/HTLV-III and normal human lymphocyte genomic DNA restriction digests. Sequences hybridizing with H9c.7 and H9c.53 were found only in DNA from infected cells, demonstrating their exogenous viral origin (Fig. 2*A*). To determine whether related sequences are associated

* Present address: Invitron, 8000 Maryland Avenue, Clayton, Missouri 63105, USA.

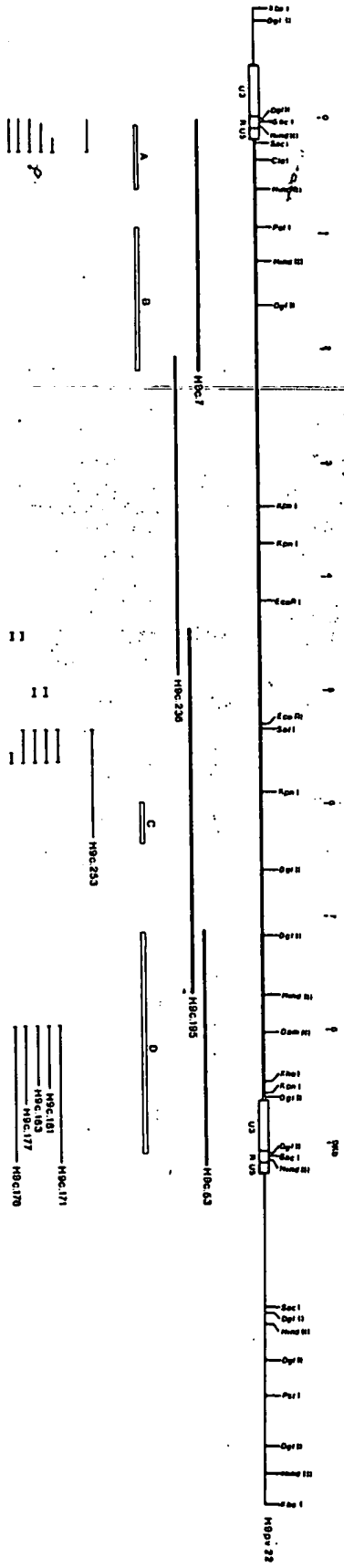


Fig. 1 Map of LAV/HTLV-III proviral and cDNA clones. Proviral sequences (bold line), flanking cellular DNA (thin line), LTR regions (boxed areas) and restriction endonuclease sites are indicated at top. Sequences are numbered from the start of viral transcription. cDNAs representing the full-length viral RNA (shown immediately below the provirus) and spliced subgenomic RNAs (at bottom of figure) are shown with the respective clone designation to the right of its map (H9pv, proviral λ clone; H9c, viral cDNA clone). The horizontal bars labelled A, B, C and D refer to fragments used as probes for the identification of spliced cDNA clones. The arrowhead at nucleotides 7,687–7,702 denotes the 16-mer primer, 5'-CTCTGTCCCACTCCAT, used to prime cDNA synthesis of clone H9c.195.

Methods. Viral cDNA clones were isolated from a cDNA library prepared from H9/HTLV-III poly(A)$^+$ RNA. Double-stranded cDNA was synthesized as described elsewhere[52] and synthetic *Eco*RI adaptors were added[53]. After the removal of excess adaptor, insertion into λgt10[54] and *in vitro* packaging, the recombinant phage were plated at a density of $\sim 5 \times 10^3$ plaques per 150-mm plate. Replica nitrocellulose filters were made from the plates[55] and the library probed with the product of uniformly ^{32}P-labelled first-strand cDNA prepared as described above from poly(A)$^+$ RNA from either H9/HTLV-III cells or the uninfected human T-cell line CEM[56]. The filters were hybridized at 42 °C in a solution containing 5×SSC, 50 mM sodium phosphate (*p*H 6.8), 0.1% sodium pyrophosphate, 5× Denhardt's solution, 0.04 g l^{-1} sonicated salmon sperm DNA, 50% formamide and 10% dextran sulphate and washed at the same temperature in 0.2×SSC, 0.1% SDS. Plaques hybridizing specifically with the H9/HTLV-III probe were plaque purified and their DNA prepared for further analysis. To isolate clones comprising the remainder of the viral RNA genome, a synthetic oligodeoxynucleotide positioned near the 5' end of clone H9c.53 was used to prime specifically cDNA synthesis. One clone obtained from the resulting cDNA library, H9c.195, extended for 3.2 kb beyond the primer location but did not overlap into the region represented by H9c.7. Rescreening of the library with the H9c.195 insert allowed the recovery of a 3.5-kb clone, H9c.236, which covered the remaining viral sequences. To isolate clones containing the complete integrated provirus, *Xba*I-digested H9/HTLV-III genomic DNA was fractionated on sucrose gradients and the resulting 10–20-kb fragments were inserted into the *Xba*I cloning site of the bacteriophage λ vector J1 (ref. 57). Twenty-six clones containing an integrated provirus were recovered from $\sim 1 \times 10^6$ recombinants screened with clones H9c.7 and H9c.53.

with LAV infection, total DNA isolated from peripheral blood lymphocytes acutely infected with LAV was hybridized with H9c.7 and H9c.53 under stringent conditions. Fragments of similar sizes were detected in *Hin*dIII, *Bgl*II and *Sst*I digests of DNA from H9/HTLV-III and LAV-infected cells (Fig. 2*A*). This result demonstrates clearly that HTLV-III and LAV correspond to the same or very closely related viruses. By contrast, H9c.7 and H9c.53 did not hybridize to cloned HTLV-I or HTLV-II provirus sequences, even in conditions of low stringency (data not shown). Significantly, the level of hybridization detected with DNA from LAV-infected lymphocytes was at least 20-fold greater than that with H9/HTLV-III DNA (Fig. 2*A*). Most of the hybridization observed with LAV-infected cell DNA migrates as a 9.5-kilobase (kb) species in *Xba*I digests (Fig. 2*A*, lane *h*) or with undigested DNA (data not shown), suggesting that this represented linear unintegrated viral DNA. By contrast, little unintegrated DNA was detected with H9/HTLV-III; instead, most of the DNA in the *Xba*I digest appeared in five or more discrete species of integrated proviruses 15–20 kb in size (Fig. 2*A*, lane *g*). Dot-blot hybridizations confirmed the presence of 5–10 copies of proviral DNA in H9/HTLV-III cells (data not shown). As no more than 5% of cells treated with LAV seemed to be infected by the virus by indirect immunofluorescence of viral antigens (data not shown), there thus seemed to be several hundred copies of unintegrated DNA present per LAV-infected cell.

Clones comprising the remainder of the viral RNA genome (H9c.236, H9c.195) were identified in a second cDNA library prepared with a specific primer (see Fig. 1). The regions represented by these four overlapping cDNA clones and their restriction maps are shown in Fig. 1. The size predicted for the full-length viral RNA genome, 9.2 kb, is consistent with the largest species observed by blot analysis of H9/HTLV-III poly(A)$^+$RNA (Fig. 2*B*). Given the possibility that the cDNA clones isolated might reflect RNA splicing events leading to the removal of small introns and therefore do not represent the entire viral genome, it was necessary to isolate molecular clones

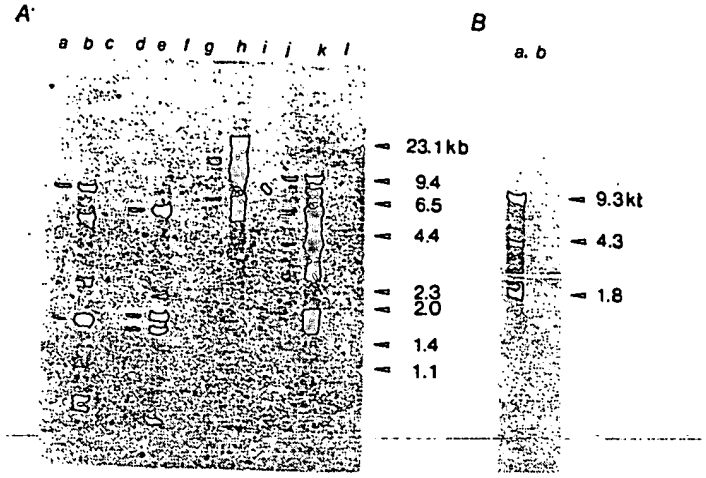


Fig. 2 Blot hybridization analysis of DNA and RNA from HTLV-III- and LAV-infected cells. *A*, Genomic DNA was isolated from the H9/HTLV-III cell line, acutely LAV-infected peripheral lymphocytes or normal blood leukocytes. Each (5 μg) was digested with the indicated restriction endonuclease, electrophoresed on a 1% agarose gel and transferred to nitrocellulose[58]. The blot was hybridized with ^{32}P-labelled probes[59] A, B and D (see Fig. 1) of equal specific activity using the hybridization and washing conditions described in Fig. 1 legend. The fragments obtained from *Hin*dIII digestion of λDNA and *Hae*III digestion of ΦX174 were used as relative molecular mass markers; these sizes are shown in kilobase pairs (kb) on the right. Lanes *a, d, g, j*, HTLV-III-infected H9 DNA; lanes *b, e, h, k*, LAV-infected lymphocyte DNA; lanes *c, f, i, l*, normal leukocyte DNA. DNA in lanes *a–c* was digested with *Hin*dIII: *d–f*, *Bgl*II; *g–i*, *Xba*I; *j–l*, *Sst*I. *B*, 1 μg poly(A)$^+$ RNA from either HTLV-III-infected H9 cells (*a*), or non-infected H9 cells (*b*) was electrophoresed on a 1% formaldehyde/agarose gel[60], transferred to nitrocellulose and hybridized to ^{32}P-labelled probe D (see Fig. 1) in hybridization and washing conditions identical to those in Fig. 1 legend. The size of each predominant RNA species was estimated from migration of single-stranded DNA markers and is indicated on the right.

of the integrated provirus. A library of size-enriched DNA was constructed to isolate proviral clones taking advantage of absence of *Xba*I sites in the provirus apparent from both the Southern blot results (Fig. 2*A*) and the map of the cDNA clones (Fig. 1). The restriction map of the integrated provirus clone selected for sequence analysis, H9pv.22, showing the proviral and adjoining cellular sequences is presented in Fig. 1.

Nucleotide sequence analysis

The complete DNA sequence of the integrated provirus is presented in Fig. 3 and compared with that determined for the four overlapping cDNA clones. The co-linearity of these sequences confirms that the cDNA clones isolated represent the unspliced viral genomic RNA. A significant degree of nucleotide heterogeneity is displayed by the proviral and cDNA sequences. Similarly, nucleotide differences are evident in the overlap regions between cDNA clones. This genomic diversity probably reflects the observation that the H9/HTLV-III cell line studied here was established by infection with material from several AIDS patients[10]. Together with the blot hybridization data indicating the presence of about five intact provirus copies (Fig. 2*A*, lane *g*), these results suggest that two or more distinct virus isolates are integrated stably in the H9/HTLV-III cell line. Despite 68 nucleotide differences between the provirus and cDNA sequences, a consistent structure nevertheless emerges for the coding potential of the virus.

The long terminal repeats (LTRs) of the provirus exhibit the structural features common to all retroviral LTRs which reflect the manner of viral replication and are essential to the mode of entry of the virus into the host genome and recognition by the cellular transcriptional apparatus[25]. The unique 5′ (U5) and 3′ (U3) regions are bordered by sequences complementary to the

Fig. 3 (Right) Complete nucleotide sequence of the HTLV-III/LAV provirus genome. The sequence shown represents the coding strand, comprising 9,213 bp extending from the RNA cap site to the polyadenylation site. The following features of the DNA sequence are indicated: the U5, R and U3 regions; splice acceptors (^{sa}) and splice donors (^{sd}); the inverted repeat located at the 3′ end of U3 and the 3′ end of U5 (IR); the tRNA^{Lys} primer binding site (PBS) and the (+)-strand initiation site (+). The polyadenylation signal (AATAAA) and Goldberg-Hogness sequence (TATAAG) are boxed. Deduced amino acid sequences of *gag*, *pol*, *P′*, *env*, and *E′* are shown above the DNA. The NH$_2$-terminus of p24^{gag} and that predicted for gp65^{env} and gp41^{env} (see text) are indicated. Nucleotide variations between cDNA and proviral isolates are below the line, and the resulting amino acid differences are above. The three nucleotides (TAA) indicated below the line at position 56–62 represent an insertion in the cDNA sequence. **Methods.** cDNA inserts were mapped with restriction endonucleases, fragments isolated and cloned into M13 vectors[61]. Single-stranded template was isolated and the sequence determined using the chain termination method[62]. Additional fragments were sequenced to determine the overlap junctions. Using the completed cDNA sequence, overlapping fragments of ~800–1,000 nucleotides were isolated from the provirus clone H9pv.22 for comparative sequence analysis.

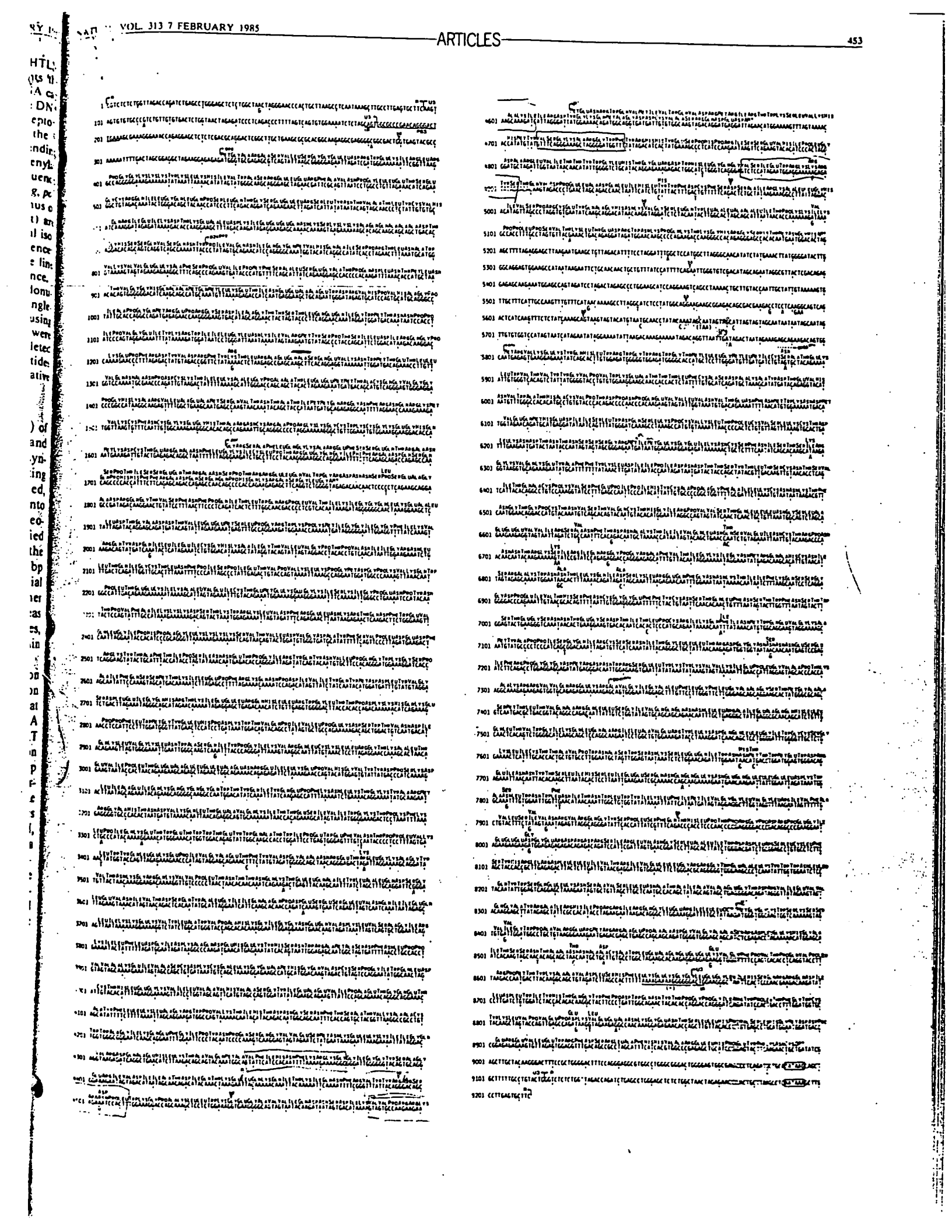
primers which copy the terminally redundant sequences (R) of the viral RNA to accomplish the series of intermolecular strand exchanges responsible for viral (−)- and (+)-strand DNA synthesis. These borders correspond to the ends of the resulting linear duplex molecule from which, in all cases so far examined, two base pairs (bp) are lost on insertion of the provirus into cellular DNA[25]. Based on this premise, the 5′ end of U3 (nucleotide 8,662) and the 3′ end of U5 (nucleotide 182) were identified from the sequence of the virus–cell DNA junctions in the H9pv.22 provirus clone (see Fig. 1). As predicted, a 23-bp sequence 3′ to the U5 boundary is complementary to a potential primer for (−)-strand synthesis, transfer RNA^{Lys} the same primer used by the mouse mammary tumour virus (MMTV)[26], whereas that 5′ to the boundary U3 consists of a stretch of 15 purines, the sequence found generally at the site of (+)-strand chain initiation[25].

In accord with the general retroviral paradigm, the integration event represented by the H9pv.22 clone reflects a duplication of host sequences at the insertion site; a short inverted repeat is found at the ends of the viral LTR. The virus–cell DNA junction sequence, TGTAGTGGGTG...CAGTGGGTGAT (viral sequences underlined), indicates a 5-bp duplication (GTGGG) of cellular DNA and an inverted repeat of 4 bp (ACTG...CAGT), two nucleotides of which are lost on insertion. In agreement with the general finding that the size of the duplication resulting from integration is a property of the virus and not the host cell, it is interesting that integration of HTLV-I, which shares OKT4 tropism with LAV/HTLV-III, results in a direct repeat of 6 bp of cellular DNA[27].

The 3′ end of the viral RNA (Fig. 3) corresponds to the site of poly(A) addition in the H9c.53 cDNA clone and is 18 nucleotides from the polyadenylation signal, AATAAA[20]. The 5′ end of the viral RNA was mapped by *in vitro* extension of a synthetic DNA primer complementary to sequences in U5. As shown in Fig. 4, most of the RNA is initiated at the G residue indicated as position 1 (Fig. 3), although heterogeneity of two or three nucleotides is observed. Based on these results, the sizes of the U3, R and U5 regions are 456, 96 and 86 nucleotides, respectively. The RNA initiation site is located 23 nucleotides from the sequence TATAAG, which conforms to the consensus Goldberg–Hogness box found characteristically in this distance and implicated in the positioning of eukaryotic transcription initiation sites[29]. A direct repeat of 10 bp is found 54 nucleotides 5′ to this sequence (nucleotides 9,013–9,022; 9,027–9,036).

The *gag* gene encodes p24

The *gag* reading frame indicated by the provirus sequence extends from nucleotides 336–1,769 and, by analogy to other retroviruses, is expected to code for a precursor polypeptide



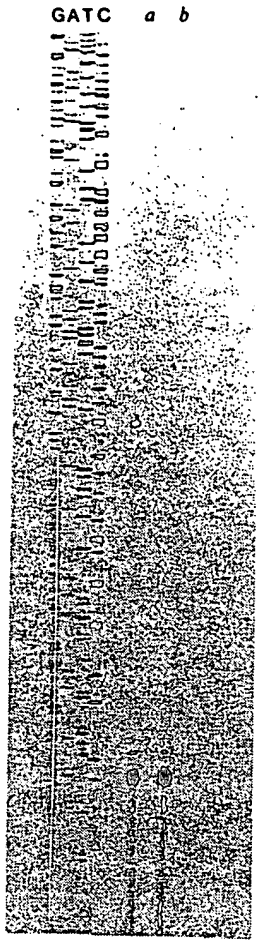


Fig. 4 Determination of the initiation site of viral transcription. Autoradiograph of a 5% polyacrylamide, 8 M urea gel. GATC, a DNA sequence ladder corresponding to the region surrounding the U3/R junction of the 5' LTR. Results of primer extension reactions are shown using as template poly(A)$^+$ RNA prepared from HTLV-III-infected (*a*) or uninfected (*b*) H9 cells.
Methods. A synthetic 16-mer complementary to the sense (+)-strand in U5 (position 129–144, see Fig. 3) was annealed to a phage M13 single-stranded template containing a portion of the (+)-strand corresponding to U5, R and part of U3. The primer was extended by incubation at 37 °C for 30 min in the presence of [α^{32}P]dCTP, [α^{32}P]dATP, dTTP, dGTP and DNA polymerase I (Klenow fragment). The product was digested with *Hin*dIII, heated at 100 °C for 5 min and fractionated on a 6% polyacrylamide gel. The labelled, 63-base-long, single-stranded fragment extending from position 144 to the *Hin*dIII cleavage site at position 81 was recovered; ~5×10^5 c.p.m. ^{32}P was annealed to 7.5 μg poly(A)$^+$ RNA in 80% formamide, 0.4 M NaCl, 40 mM PIPES (*p*H 6.5), 1 mM EDTA at 52 °C for 3 h. After primer extension with avian myeloblastosis virus reverse transcriptase, the sample was treated with 100 mM NaOH at 70 °C for 45 min and loaded onto a 5% polyacrylamide, 8 M urea gel. The DNA sequence used as marker was generated by using the above 16-mer as the primer in dideoxy method sequencing reactions on the same M13 template used to prepare the 63-bp fragment, therefore the primer extension products co-migrate with the corresponding nucleotide sequence.

which is post-translationally cleaved to give the internal structural proteins of the virus. Translation of this reading frame begins at the 5'-proximal ATG triplet of the viral genomic RNA and would lead to the synthesis of a 478 amino acid polypeptide consistent with the relative molecular mass (M_r) 53–55,000 of the *gag*-encoded polyprotein (Pr54^{gag}) detected recently in H9/HTLV-III cells[30]. A good candidate for the major virion core protein encoded by this precursor is a 24,000 M_r virus-associated protein (p24) recognized by sera from AIDS and ARC patients[7–9,15,16] (see Fig. 6). Recently, the N-terminal amino acid sequence of p24 has been determined (J. Bell and C.V.B., unpublished observations) and the 17-residue sequence obtained for the purified protein matches exactly that beginning with the proline residue found at position 732 (Fig. 3).

The p24^{gag} N-terminal cleavage thus identified predicts a 13-amino acid protein from the N-terminus of the *gag*-encoded polyprotein, containing a single potential asparagine-linked N-glycosylation site. Interestingly, this cleavage recognition site (the aromatic amino acid proline) is analogous to that found in the precursor for Moloney murine leukaemia virus (Mo-MuLV)[31], suggesting that processing occurs by a viral or cellular protease with similar specificity.

Although the proteolytic cleavage responsible for generating the C-terminus of p24^{gag} has not yet been defined, the presence of ~130 amino acids beyond the sequence sufficient to encode p24^{gag} indicates that a third protein is encoded by Pr54^{gag}. A direct repeat of 36 nucleotides resulting in a C-terminal-duplication of 11 amino acids (positions 1,676–1,747) is obvious in this region. The sequence of this protein shows significant conservation of cysteine residues with p12^{gag} of Rous Sarcoma Virus (RSV), p11^{gag} of HTLV-I and p10^{gag} of Mo-MuLV[27,31,32]. Common to each protein are three cysteine residues separated by two and nine amino acids, respectively; this structure is found twice in the HTLV-III, RSV and HTLV-I proteins and once in Mo-MuLV. The p12^{gag} of RSV, a protein rich in basic residues, is the major component of the virion ribonucleoprotein complex, binding nonspecifically to many sites on the viral RNA[33,34]. Similarly, the HTLV-III protein is highly basic, containing 29 lysine and arginine residues. These striking similarities between otherwise highly-diverged proteins suggests that this C-terminal *gag*-encoded protein constitutes the core ribonucleoprotein.

The *pol* region

The product of the *pol* gene is encoded by the second large reading frame of the virus located between nucleotides 1,639–4,674. The predicted amino acid sequence shows significant homology to the *pol* gene products of RSV, HTLV-I, and Mo-MuLV (~20–30%)[27,31,32], confirming the assignment of this reading frame to the *pol* gene. Despite the conservation of amino acid structure, however, there is no significant corresponding nucleotide homology with RSV, HTLV-I or Mo-MuLV, in contrast to earlier reports demonstrating cross-hybridization between HTLV-III and HTLV-I in this region[35,36]. The reverse transcriptase of several avian and mammalian retroviruses is translated initially as a *gag–pol*-encoded polyprotein, an event for which either the suppression of an in-frame amber codon (Mo-MuLV) or the removal of a short interval by RNA splicing (RSV) has been suggested as a mechanism for joining the reading frames of the *gag* and *pol* genes[31,32]. The *gag* and *pol* genes of HTLV-III are also situated in different reading frames (Fig. 3), suggesting a similar requirement for splicing as a mechanism of *pol* expression. The first ATG codon encountered in the *pol* gene is at position 1,939; however, the potential reading frame begins at position 1,639, overlapping the *gag* gene by 130 nucleotides. Thus, the possibility exists for shared amino acid sequences between the C-terminus of the *gag*-encoded ribonucleoprotein and the N-terminus of reverse transcriptase.

There exists a third reading frame, designated *P'*, between nucleotides 4,589–5,197, containing ATG triplets at positions 4,622, 4,643, 4,667 and 4,706 and could thus encode a protein of ~192 amino acids. The potential for generating spliced transcripts containing *P'* sequences suggests that this reading frame is used (see below). Beyond *P'* there is an intercistronic region of 602 nucleotides with frequent stop codons in all three reading frames. This contrasts with the compact organization of Mo-MuLV and HTLV-I that have overlapping *pol* and *env* genes[27,31].

The *env* gene

The primary translation product of the *env* gene of retroviruses is a large glycosylated precursor synthesized on the rough endoplasmic reticulum, which is processed to produce a larger N-terminal glycoprotein bearing the host range determinants and a smaller hydrophobic protein serving as the membrane anchor by virtue of a transmembrane domain. The *env* gene product

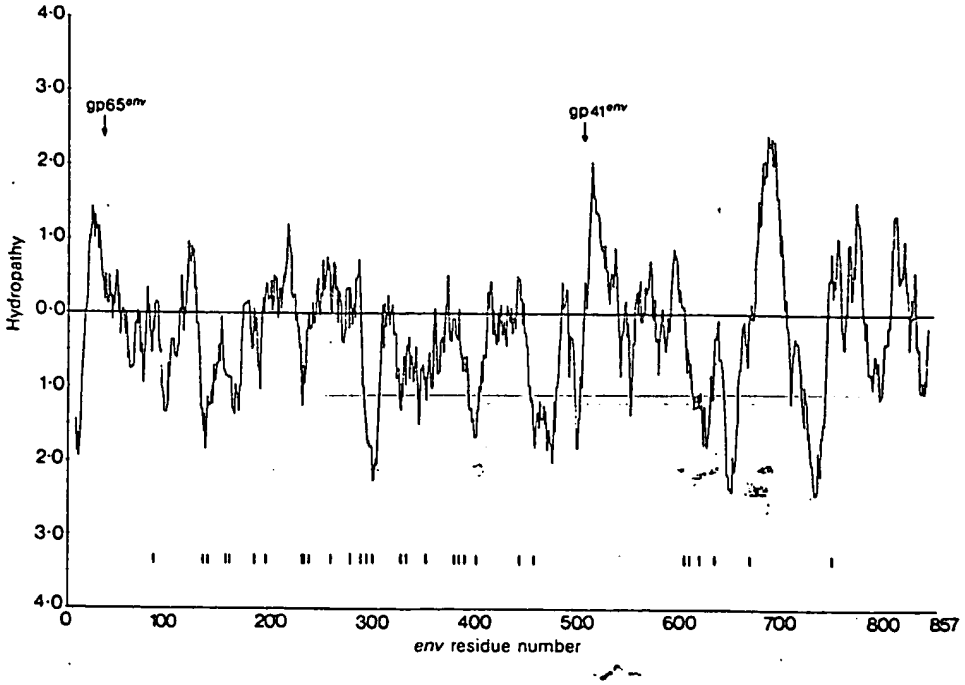


Fig. 5 Hydropathy plot for the *env* gene product. Hydrophobic areas appear above the midline, and hydrophilic areas below. Potential sites of asparagine-linked glycosylation are indicated by vertical lines below the plot. The NH_2-termini of the putative mature $gp65^{env}$ and $gp41^{env}$ are indicated. The region corresponding to $gp65^{env}$ spans ~ 480 amino acids beginning after a hydrophobic leader and extending to the conserved processing site Arg-N-(Arg/Lys)-Arg—which marks the NH_2-terminus of $gp41^{env}$. The $gp41^{env}$ region consists of 345 amino acids containing two extended hydrophobic domains. Hydropathy was calculated by the method of Kyte and Doolittle[63] for each overlapping segment of 15 amino acids.

of HTLV-III is encoded by the large open reading frame at nucleotides 5,782–8,370. Analysis of a cDNA clone representing the *env* mRNA (see below) suggests that translation of the *env* region is initiated at the ATG codon located near the beginning of this reading frame at position 5,803. This assignment predicts the synthesis of an 856-residue envelope precursor protein containing 30 potential sites of asparagine glycosylation located principally in the first half of the molecule (Fig. 5). Taking into account the addition of carbohydrate, the predicted size of the *env* precursor approximates gp120, a 120,000 M_r glycoprotein detected recently in intracellularly labelled H9/HTLV-III cells[30].

Although N-terminal sequence information is presently unavailable for the envelope proteins of HTLV-III and there is no discernible amino acid homology with the *env* gene products of RSV, Mo-MuLV, HTLV-I or MMTV[27,31,32,37], certain features of the sequence allow the prediction of the processing of the precursor molecule. Three stretches of hydrophobic and nonpolar residues are present at positions 5,851–5,886 (12 residues), 7,336–7,419 (28 residues) and 7,852–7,917 (22 residues) (Fig. 5). The first stretch is located at the N-terminus of the precursor and is flanked by charged residues, suggesting a role as the signal sequence responsible for directing the protein to the cell surface[38]. The size and position of the other two hydrophobic regions suggest that they are located in the transmembrane protein of the HTLV-III envelope. The transmembrane envelope proteins of RSV, Mo-MuLV, MMTV and HTLV-I similarly display two hydrophobic stretches of 20–30 amino acids separated by ~150–200 residues[27,31,32,37]. In each case, maturation of the envelope polyprotein involves cleavage after a conserved sequence of basic residues, Arg-N-(Arg/Lys)-Arg, immediately preceding the first hydrophobic stretch. The presence of this sequence at the corresponding position of the HTLV-III *env* gene product suggests that $gp120^{env}$ is cleaved into a N-terminal protein of ~480 amino acids and a transmembrane protein of 345 amino acids with 24 and 6 potential asparagine-linked glycosylation sites, respectively (Fig. 5).

This assignment for the major envelope glycoprotein and transmembrane protein is supported by serological evidence. We analysed HTLV-III virion proteins by immunoblotting with antisera from individuals infected with the virus. With extensively diluted sera from infected individuals, we detect a predominant specis of M_r 65,000 in addition to $p24^{gag}$ (Fig. 6*A–E*), which does not react with undiluted sera from normal individuals (Fig. 6*F*) suggesting that p65, like $p24^{gag}$, is a major structural constituent of the virion. The size of p65 is consistent with the predicted size of the major envelope glycoprotein. Previous studies have indicated the presence of a p65 in HTLV-III virion preparations and in H9/HTLV-III cells[15,16,39], but have consistently detected far greater amounts of a 41,000 M_r glycoprotein[16,17] (also detected by the sera analysed in Fig. 6). In light of the present evidence, it seems that gp41 may represent the transmembrane protein rather than the major glycoprotein. The apparent difference in the ability of patient sera to detect p65 rather than gp41 in these studies may reflect the method of virus preparation.

Beyond the *env* gene there is a fifth open reading frame, designated *E'*, between nucleotides 8,347–8,992, extending into the U3 region of LTR. This novel reading frame can encode a protein of 206 amino acid residues, beginning at the ATG triplet at nucleotide 8,370, which is unrelated to the X genes of HTLV-I or HTLV-II (refs 27, 40, 41).

Transcription analysis

The synthesis of proteins encoded at internal sites in the retroviral genome is accomplished by two mechanisms; the post-translational cleavage of polyprotein precursors and for genes located in the 3' half of the genome (the *env* gene and transforming genes of several acute transforming viruses), by the expression of spliced subgenomic messenger RNAs. Northern analysis of H9/HTLV-III RNA revealed three prevalent viral specific RNAs of 9.3, 4.3 and 1.8 kb (Fig. 2*B*), suggesting the expression of three major virus-encoded primary translation products. To determine whether the 4.3- and 1.8-kb subgenomic RNAs could account for the synthesis of the HTLV-III *env* and *E'* gene products, viral transcription was analysed with hybridization probes specific for the detection of cDNA clones representing spliced transcripts. Evidence from several retroviruses indicates that sequences from the 5' end of the viral genome are found in each of the major viral RNAs[42–44]. Therefore, we sought cDNA clones that would hybridize to probes derived from the 5' end of the virus (probe A) as well as the *env* (probes C, D) or *E'* regions (probe D), but not with sequences corresponding to the *gag-pol* region (probe B) (Fig. 1).

This screening strategy allowed the identification of two distinct classes of subgenomic mRNA clones. The structures of these transcripts (obtained by DNA sequence analysis) are presented in Fig. 1. The first class of spliced mRNAs, typified by clone H9c.253, is comparable to the RNA species encoding the envelope proteins of other retroviruses[45]. Transcription of this class of mRNA is initiated in the 5' LTR, probably extends to the polyadenylation site in the 3' LTR and reflects the removal of a large intron between nucleotides 289–5,359 containing the

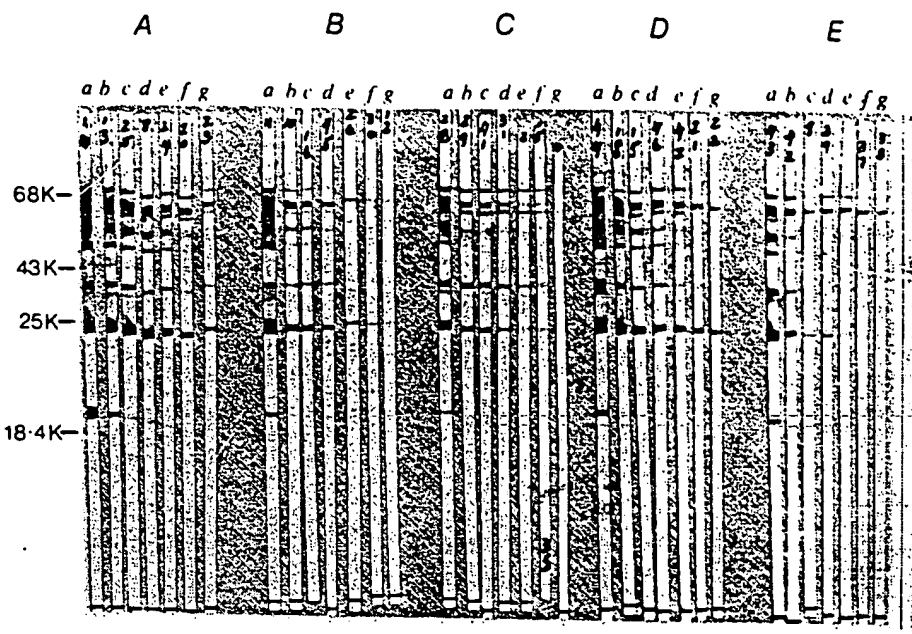


Fig. 6 Immunoblot detection of viral antigens. Virion-specific proteins were detected by sera from two healthy homosexual men (*A*, *D*), plasma from two other healthy homosexual men (*B*, *C*), serum from a sexual partner of an AIDS patient (*E*), but not by undiluted serum from a healthy control individual (*F*). Serum samples were undiluted (*a*), diluted 1:10 (*b*), 1:20 (*c*), 1:40 (*d*), 1:80 (*e*), 1:160 (*f*), 1:320 (*g*). At the highest sample dilutions, only p24 and p65 are detectable.

Methods. HTLV-III-infected H9 cells were grown in RPMI 1640 supplemented with 5% fetal bovine serum. Virus particles were concentrated from culture fluids by ultrafiltration and purified by banding on a 20–60% sucrose gradient in an SW 27 rotor centrifuged for 16 h at 27,000 r.p.m. Fractions between 30–40% were diluted and the viral particles pelleted by ultracentrifugation. Viral proteins were separated on a 12% SDS-polyacrylamide gel and transferred to nitrocellulose. The blot was cut into strips and incubated overnight at 25 °C with the patient sample. After washing, virus-specific protein bands were visualized by incubation with biotinylated goat anti-human IgG and subsequent incubation with avidin-horseradish peroxidase. M_r standards of 68, 43, 25 and 18.4 ($\times 10^3$) are shown.

gag, *pol* and *P'* genes. The size predicted for this RNA, 4,144 nucleotides plus poly(A), corresponds to the 4.3-kb species detected by Northern analysis (Fig. 2*B*). The putative *env* initiation codon is preceded by three ATG triplets at positions 5,412, 5,551 and 5,643; however, each is followed by a nearby in-frame stop codon located 216, 81 and 243 nucleotides 3', respectively. The selection of an ATG codon other than the most 5' proximal ATG triplet for translation initiation has been previously noted for *gag* and *env* expression in Mo-MuLV and RSV[31,32].

The second class of spliced mRNAs identified hybridized with probe D but not probe C, indicating that they did not contain most of the *env* gene. This class of mRNAs reflected a variety of splicing events, involving the removal of two or more introns (Fig. 1) with the generation of an RNA species similar in size (~1.8 kb) to the smaller subgenomic RNA seen by blot analysis (Fig. 2*B*). Each member of this class was formed from the 5' leader (exon 1) and sequences corresponding to the first 268 nucleotides of *env* leader (exon 4). In every case, the latter sequences were joined directly to the splice acceptor at position 7,957 in the C-terminal coding region of *env* (exon 7). In addition, several clones possessed an additional untranslated leader exon of 50 nucleotides (exon 2) (H9c.181, H9c.183) or 74 nucleotides (exon 3) (H9c.176, H9c.177) derived from the *pol* and *P'* regions, respectively. Table 1 summarizes the size and location of each exon and compares their donor and acceptor sequences. As a consequence of the splicing event which deletes most of the *env* gene, the ATG triplet at position 5,643 is removed whereas the distance between the ATG triplets at positions 5,412 and 5,551 and the next in-frame stop codons increases to 258 and 348 nucleotides, respectively. Neither of these reading frames connects with the sequences that encode the C-terminus of the *env* polyprotein. One of the clones (H9c.176) uses an alternative splice acceptor in exon 4, resulting in the juxtaposition of exon 2 with the last 69 nucleotides of the *env* leader (exon 5) and the removal of the ATG codons at 5,412 and 5,551. Significantly, the members of this diverse class of spliced RNAs are all related by their ability to direct the synthesis of the putative *E'* gene product.

Discussion

Although a definitive interpretation of the HTLV-III sequence awaits the demonstration of biological activity for our virus clones, the concordance of provirus and cDNA sequences corresponding to apparently distinct virus isolates suggests that they portray accurately the structural features of the viral genome. Despite a gene organization in many aspects similar to other retroviruses, the HTLV-III genome seems to be entirely unrelated by nucleotide homology to previously characterized retroviral sequences[27,31,32,37], including HTLV-I and HTLV-II which display a similar tropism for the OKT4$^+$ T-cell subset[46]. The

Table 1 Summary of LAV/HTLV-III exon and splice junctions

Exon	Location	Length (nucleotides)	Splice acceptor	Splice donor
1	1–289	289	–	... GACTG | GTGAGTAC
2	4,494–4,543	50	TTTCGGGTTTATTACAG | GGA	... GAAAG | GTGAGGG
3	4,971–5,044	74	CTTTGACTGTTTTTCAG | ACT	... ACAAG | GTAGGATC
4	5,359–5,626	268	TGTTTATCCATTTTCAG | AAT	... AAGCA | GTAAGTAG
5	5,558–5,626	69	GGCATCTCCTATGGCAG | GAA	... AAGCA | GTAAGTAG
6	5,359–9,213	3,855	TGTTTATCCATTTTCAG | AAT	–
7	7,957–9,213	1,257	CACCATTATCGTTTCAG | ACC	–

Each row lists the known exons (see Fig. 7), the nucleotide positions comprising each exon (see Fig. 3), and total exon length. The DNA sequence adjacent to the 5' (acceptor) and 3' (donor) borders at each exon are also shown. Intron sequences are to the left of the vertical line in the acceptor column, and to the right of the vertical line in the donor column. These sequences were obtained by comparison of the complete proviral sequence with those of several cloned cDNAs representing spliced subgenomic mRNAs, shown in Fig. 1.

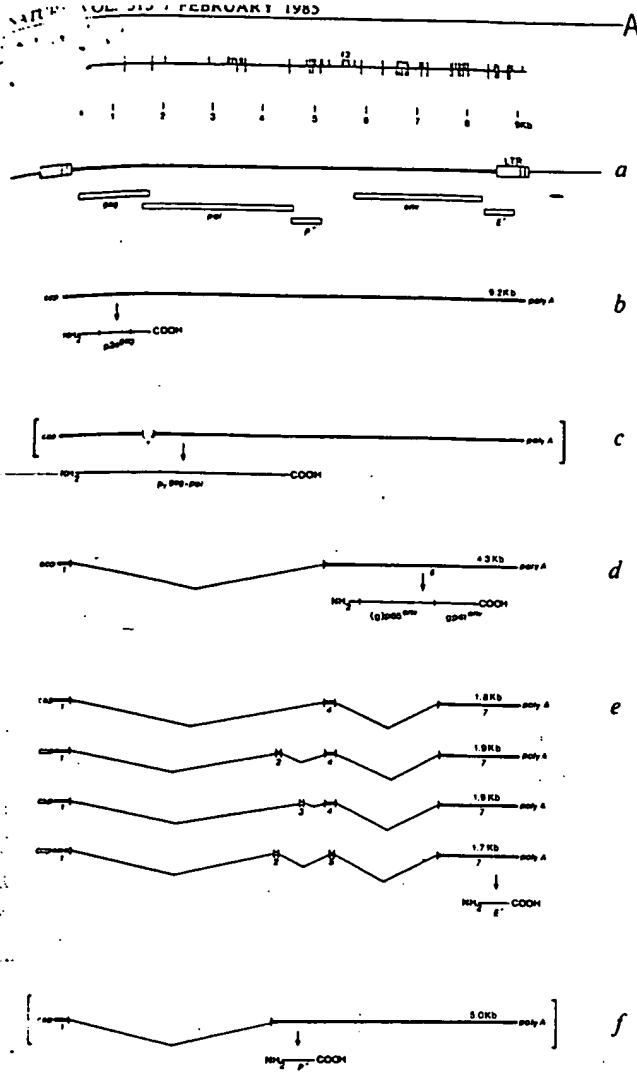


Fig. 7 Summary of viral gene organization and expression. *a*, Translational reading frames for the viral genes are indicated below a representation of the integrated proviral genome. Vertical marks above the line denote the location of nucleotide differences between the proviral (H9pv.22) and cDNA clones (H9c.7, H9c.236, H9c.195, H9c.53; vertical marks below the line indicate the resulting amino acid substitutions. *b*, Synthesis and subsequent processing of the *gag* precursor from full-length viral RNA. *c*, Hypothetical splicing pathway similar to that proposed for RSV[32] which would allow for synthesis of a *gag-pol* fusion protein. *d*, Synthesis and processing of the *env* polyprotein from the 4.3-kb spliced, subgenomic mRNA. *e*, Predicted synthesis of *E'* gene product from the 1.8-kb class of spliced, subgenomic mRNAs. *f*, Hypothetical subgenomic RNA joining the 5' leader with exon 2 without subsequent splicing that could encode the predicted *P'* gene product.

most significant protein homology detected between the gene products of HTLV-III and other retroviruses is in the *pol* gene (~20–30%); however, a corresponding degree of nucleotide relatedness is not found, which is inconsistent with reports of homology to HTLV-I in the *gag-pol* region based on hybridization data[35,36]. These results and the observation that the virus is morphologically distinct from the type C viruses[8,13,15] lead us to propose that LAV/HTLV-III is a member of a novel class of retroviruses, perhaps including the equine infectious anaemia virus, which has a similar morphology and a serologically-related major core protein[8].

Figure 7 summarizes the gene organization, transcriptional processing and polyprotein maturation pathways predicted for the virus. The *gag* gene product is a precursor of M_r 54,000 which, as found for other retroviruses, seems to be synthesized from genome-size viral mRNA (Fig. 7*b*). $Pr54^{gag}$ is cleaved at two positions to generate the major core structural protein, p24, and M_r ~15,000 N-terminal polypeptide and a structurally-conserved C-terminal ribonucleoprotein of M_r ~15,000. The *pol* and *gag* genes are encoded by different reading frames; translation of a *gag-pol*-encoded polyprotein would thus require joining of the reading frames by a splicing event (Fig. 7*c*).

Characterization of a cDNA clone which probably represents the 4.3-kb major viral subgenomic RNA has allowed the identification of the transcript capable of directing the stoichiometric synthesis of the *env* gene products. The removal of a single intron containing the *gag-pol* region accomplishes the joining of a 289-nucleotide leader from the 5' end of the genome with a second untranslated leader situated within an intercistronic region between *pol* and *env* (Fig. 7*d*). Translation of this spliced subgenomic RNA would lead to synthesis of an 856 amino acid envelope polyprotein with a hydrophobic signal leader which would direct the precursor to the cell membrane and allow initiation of the viral envelope formation. Evidence of a conserved cleavage recognition site preceding an extended hydrophobic domain suggests processing of $Pr120^{env}$ generates the major envelope and transmembrane glycoproteins. The sizes thus predicted for $gp65^{env}$ and $gp41^{env}$ are ~480 and 345 residues, respectively, in agreement with the observed electrophoretic mobilities of two major virion structural proteins. Demonstration of the glycoprotein nature of p65 or direct sequence analysis of these polypeptides will be required to affirm these assignments.

Expression of the novel *E'* gene product is indicated by a class of cDNA clones corresponding to several related but structurally heterogeneous spliced subgenomic RNAs of ~1.8 kb (Fig. 7*e*). The salient feature of this class is a tripartite structure consisting of the 5' leader, the intercistronic leader and 1.3 kb of RNA from the 3' end of the genome. Translation of the *E'* reading frame would result in the synthesis of a 206-residue protein lacking homology with the *X* gene products encoded by the 3' regions of HTLV-I and HTLV-II. Differentially-spliced mRNAs capable of encoding the *E'* gene product are generated by the addition of either of two additional untranslated leaders or by the use of an alternative splice acceptor in the intercistronic leader sequence. Although the existence of *E'* is unprecedented in other retroviral genomes, it is interesting to speculate that it acts as a virus-specific transcription factor, a function ascribed recently to the HTLV-I and HTLV-II *X* gene products[47,48]. The multiplicity of splicing patterns also provides a rationale for the generation of a subgenomic mRNA directing the synthesis of the gene product encoded by the *P'* reading frame (Fig. 7*f*). In the examples already described, the joining of the 5' leader to the small exon in the *pol* gene (exon 2) is accompanied always by further splicing at the donor site located 50 nucleotides 3'. If further splicing at this donor did not occur, a subgenomic RNA ~870 nucleotides longer than the *env* mRNA would be produced having as its primary translation product a 192 amino acid protein specified by the *P'* reading frame. Alternatively, the removal of a small intron between the *pol* and *P'* reading frames could result in the synthesis of a *gag-pol-P'* precursor peptide. Further investigation will be required to establish the identity the *E'* and *P'* gene products and their roles in viral reproduction and pathogenesis.

Following retroviral infection, the major product of viral DNA synthesis in the cytoplasm is a linear, double-stranded molecule of genome size with a complete copy of the LTR at each end[25]. Studies with several cytopathic retroviruses, notably spleen necrosis virus and the cytopathic subgroups of avian leukosis virus, have revealed a correlation between transient cell killing during acute infection and the transient accumulation of 100–200 copies of linear unintegrated viral DNA[49–51]. Both transient cell killing and transient accumulation of linear unintegrated DNA require the spread of virus and superinfection of the infected cell population. Once a chronic state is established, the level of linear unintegrated DNA in the surviving cells decreases ~100-fold and is accompanied by the stable integration of several copies of the provirus. It is therefore interesting that similar preliminary observations have been made for LAV/HTLV-III infection of T lymphocytes. Southern blot analysis of DNA from

the chronically infected H9/HTLV-III T-cell line reveals 5–10 copies of integrated provirus, although there is little evidence for the persistence of more than a small amount of linear unintegrated DNA. By contrast, acute infection of human peripheral lymphocyte cultures with LAV leads to the apparent accumulation of >400 copies per cell of linear unintegrated DNA. Although H9/HTLV-III cells are relatively resistant to the cytopathic effects of HTLV-III[10], the rapid disappearance of virus-producing cells in infected primary lymphocyte cultures suggests that LAV/HTLV-III has a severe cytopathic effect upon the natural host target cell ($OKT4^+$)[8,11,14]. The level of linear unintegrated DNA in H9/HTLV-III cells and infected lymphocyte cultures thus seems to correlate with the relative cytopathic effects of the virus on these cells.

The H9/HTLV-III cell line was established by infection with material from several AIDS patients[10] and therefore may contain sequences corresponding to different viral isolates. The distribution of nucleotide differences between the proviral and cDNA sequences described here must be interpreted with caution as we do not know the exact relationship of a particular clone to any of the multiple provirus copies present in the H9/HTLV-III cell line (Fig. 2*A*). Nevertheless, the independent origin of some of the clones is suggested by the large number of nucleotide differences present (Fig. 7*a*). Particularly evident is a high degree of polymorphism in the *env* coding region (24 of 2,568 nucleotides are different) and the intercistronic region (15/602). Eighteen of the changes in *env* lead to amino acid substitutions, mostly nonconservative. Two changes result in new potential sites of asparagine-linked glycosylation while one change removes such a site and therefore may be especially likely to have significant effects on protein antigenicity. Should the proclivity for change with the HTLV-III envelope seen here reflect a high rate of mutation *in vivo*, this may have significant implications for the ability of the immune system to respond effectively to the virus as well as for the design of vaccines.

We thank Dr Jerome Groopman for providing serum samples, Dr Larry Arthur for providing the H9/HTLV-III$_B$ cell line established by Dr Robert Gallo; Dr John Bell for communicating unpublished data; Parkash Jurhani and Peter Ng for DNA synthesis, Dr Frances Boches for assistance with the immune blotting experiments; Carol Morita for preparation of the figures and Rebecca Cazares for help in preparation of the manuscript. We also thank Drs Jack Obijeski, David Goeddel, Peter Seeburg and Rey Gomez for support and encouragement and Drs Arthur Levinson, Dennis Kleid and David Martin for comments on the manuscript.

Received 27 December 1984; accepted 22 January 1985.

1. Gottlieb, M. *et al. New Engl. J. Med.* 305, 1425–1431 (1981).
2. Masur, H. *et al. New Engl. J. Med.* 305, 1431–1438 (1981).
3. Siegal, F. *et al. New Engl. J. Med.* 305, 1439–1444 (1981).
4. Seligman, M. *et al. New Engl. J. Med.* 311, 1286–1292 (1984).
5. Lane, H., Masur, H., Edgar, G., Whalen, G. & Fauci, A. *New Engl. J. Med.* 309, 453–458 (1983).
6. Ziegler, J. *et al. New Engl. J. Med.* 311, 565–570 (1984).
7. Barre-Sinoussi, F. *et al. Science* 220, 868 (1983).
8. Montagnier, L. *et al. Human T-Cell Leukemia Viruses*, 363–379 (Cold Spring Harbor Laboratory, New York, 1984).
9. Vilmer, E. *et al. Lancet* i, 753–757 (1984).
10. Popovic, M., Sarngadharan, M., Read, E. & Gallo, R. *Science* 224, 497–500 (1984).
11. Gallo, R. *et al. Science* 224, 500–503 (1984).
12. Feorino, P. *et al. Science* 225, 69–72 (1984).
13. Levy, J. *et al. Science* 225, 840–842 (1984).
14. Klatzmann, D. *et al. Science* 225, 59–63 (1984).
15. Schupbach, J. *et al. Science* 224, 503–505 (1984).
16. Sarngadharan, M., Popovic, M., Bruch, L., Schupbach, J. & Gallo, R. *Science* 224, 505–508 (1984).
17. Safai, B. *et al. Lancet* i, 1438–1440 (1984).
18. Brun-Vezinet, F. *et al. Lancet* i, 1253–1256 (1984).
19. Brun-Vezinet, F. *et al. Science* 226, 453–456 (1984).
20. Goedert, J. *et al. Lancet* ii, 711–715 (1984).
21. Lawrence, J. *et al. New Engl. J. Med.* 311, 1269–1273 (1984).
22. Schwartz, D., Zamecnik, P. & Weith, H. *Proc. natn. Acad. Sci. U.S.A.* 74, 994–998 (1977).
23. Haseltine, W., Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* 74, 989–993 (1977).
24. Shine, J. *et al. Proc. natn. Acad. Sci. U.S.A.* 74, 1473–1477 (1977).
25. Varmus, H. *Science* 216, 812–820 (1982).
26. Peters, G. & Glover, C. *J. Virol.* 35, 31–40 (1980).
27. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* 80, 3618–3622 (1983).
28. Proudfoot, N. & Brownlee, G. *Nature* 252, 359–362 (1974).
29. Grosschedl, R. & Birnstiel, M. *Proc. natn. Acad. Sci. U.S.A.* 77, 1432–1436 (1980).
30. Kitchen, L. *et al. Nature* 312, 367–369 (1984).
31. Shinnick, T., Lerner, R. & Sutcliffe, J. *Nature* 293, 543–548 (1981).
32. Schwartz, D., Tizard, R. & Gilbert, W. *Cell* 32, 853–869 (1983).
33. Bolognesi, D., Luftig, R. & Schafer, J. *Virology* 56, 549–564 (1973).
34. Smith, B. & Bailey, J. *Nucleic Acids Res.* 7, 2055–2072 (1979).
35. Arya, S. *et al. Science* 225, 927–930 (1984).
36. Hahn, B. *et al. Nature* 312, 166–170 (1984).
37. Majors, J. & Varmus, H. *J. Virol.* 47, 495–504 (1983).
38. Perlman, D. & Halvorson, H. *J. molec. Biol.* 167, 394–409 (1983).
39. Montagnier, L. L. *et al. Science* 225, 63–66 (1984).
40. Haseltine, W. *et al. Science* 225, 419–421 (1984).
41. Shimotohno, K. *et al. Proc. natn. Acad. Sci. U.S.A.* 81, 6657–6661 (1984).
42. Weiss, S., Varmus, H. & Bishop, M. *Cell* 12, 983–992 (1977).
43. Mellon, P. & Duesberg, P. *Nature* 270, 631–634 (1977).
44. Cordell, B., Weiss, S., Varmus, H. & Bishop, M. *Cell* 15, 79–91 (1978).
45. Weiss, R., Teich, N., Varmus, H. & Coffin, J. *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1982).
46. Popovic, M. *et al. Science* 219, 856 (1983).
47. Sodrowski, J., Rosen, C. & Haseltine, W. *Science* 225, 381–385 (1984).
48. Chen, I., McLaughlin, J. & Golde, D. *Nature* 309, 276–279 (1984).
49. Keshet, E. & Temin, H. *J. Virol.* 31, 376–388 (1979).
50. Weller, S., Joy, A. & Temin, H. *J. Virol.* 33, 494–506 (1980).
51. Temin, H., Keshet, E. & Weller, S. *Cold Spring Harbor Symp. quant. Biol.* 44, 773–777 (1980).
52. Capon, D. *et al. Nature* 304, 507–513 (1983).
53. Wood, W. *et al. Nature* 312, 330–337 (1984).
54. Hyunh, T., Young, R. & Davis, R. in *Practical Approaches to Biochemistry* (ed. Groves, D.) (IRL, Oxford, in the press).
55. Benton, W. & Davis, R. *Science* 196, 180–182 (1977).
56. Foley, G. *et al. Cancer* 18, 522–529 (1965).
57. Mullins, J., Brody, D., Binari, R. & Cotter, S. *Nature* 308, 856–858 (1984).
58. Southern, E. *J. molec. Biol.* 98, 503–517 (1975).
59. Taylor, J., Illmensee, R. & Summer, S. *Biochim. biophys. Acta* 442, 324–330 (1976).
60. Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning* (Cold Spring Harbor Laboratory, New York, 1982).
61. Messing, J., Crea, R. & Seeburg, P. *Nucleic Acids Res.* 9, 309–321 (1981).
62. Sanger, F. *et al. J. molec. Biol.* 143, 161–178 (1980).
63. Kyte, J. & Doolittle, R. *J. molec. Biol.* 157, 105–132 (1982).

A *trans*-acting factor is responsible for the simian virus 40 enhancer activity *in vitro*

Paolo Sassone-Corsi, Alan Wildeman & Pierre Chambon

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Faculté de Médecine, 11 Rue Humann, 67085 Strasbourg Cédex, France

Stimulation of in vitro *transcription by the simian virus 40 enhancer involves a rapid and stable binding of a trans-acting factor with both the 5'- and 3'-domains of the enhancer sequence. The enhancer factor, which differs from other types of transcriptional factors, can interact with other enhancer elements.*

REGULATION of gene expression at the level of transcription is probably an important control mechanism during development and in the terminally differentiated cells of eukaryotic organisms. This control may result from the interaction between specific DNA sequences, regulatory proteins and the transcriptional machinery[1]. The promoter DNA sequences involved in the control of transcription initiation in eukaryotic protein-coding genes are composed of several elements: the mRNA start-site, the TATA box sequence and one or several upstream elements, located generally in ~110 base pairs (bp) upstream

VIROLOGY 185, 661–672 (1991)

Changes in Growth Properties on Passage in Tissue Culture of Viruses Derived from Infectious Molecular Clones of HIV-1_{LAI}, HIV-1_{MAL}, and HIV-1_{ELI}

KEITH PEDEN,*,†,[1] MICHAEL EMERMAN,*,[2] AND LUC MONTAGNIER*

*Unité d'Oncologie Virale, Institut Pasteur, 28 rue du Dr Roux, Paris 75015, France, and †Laboratory of Molecular Microbiology, NIAID, NIH, Bethesda, Maryland 20892

Received July 8, 1991; accepted August 22, 1991

The construction and preliminary biological characterization of three molecular clones of human immunodeficiency virus type 1 (HIV-1) are reported: HIV-1_{LAI} from a French man with AIDS, HIV-1_{MAL} from a Zairian boy with ARC, and HIV-1_{ELI} from a Zairian woman with AIDS. All three sequences were found to code for infectious viruses. Both the host range and the kinetics of infection in CD4$^+$ cells were different for the three viruses. Virus derived from each molecular clone was infectious on peripheral blood mononuclear cells (PBMC), although LAI and ELI displayed more rapid growth kinetics than MAL. The viruses had different tropisms and growth kinetics in six cell lines. LAI was infectious in all of the cell lines and produced high levels of reverse transcriptase activity. MAL and ELI had more restricted tropisms: MAL could only replicate on SupT1, whereas ELI grew on Jurkat and MT-4, was delayed on CEM and H9, and was unable to infect U937 cells. In addition, we observed that both the replicative capacity and the cell tropism of viruses could change after passage through some established cell lines. These results suggest that the genotypes of some viruses *in vitro* are not stable and that selection for growth can cause the fairly rapid appearance of variants with increased growth potential. © 1991 Academic Press, Inc.

INTRODUCTION

The virus causing the majority of acquired immune deficiency syndrome (AIDS) cases throughout the world remains the human immunodeficiency virus type 1 (HIV-1). The determination of the nucleotide sequence of the LAI strain of HIV-1 and its derivatives revealed a genome structure larger and more complex than that hitherto described for other retroviruses (Wain-Hobson *et al.*, 1985; Ratner *et al.*, 1985; Muesing *et al.*, 1985). Sequencing additional viral genomes, both from North America (Sanchez-Pescador *et al.*, 1985) and central Africa (Alizon *et al.*, 1986), revealed a remarkable sequence diversity among different HIV-1 isolates (Sonigo *et al.*, 1985). HIV-1 has nine open reading frames (ORFs) divided into three structural and replication genes (*gag, pol,* and *env*), two regulatory genes (*tat* and *rev*), and four accessory genes (*vif, vpr, vpu,* and *nef*); the functions of the products of this last class are poorly understood.

HIV-1_{LAI} was isolated from a French homosexual male with AIDS and Kaposi sarcoma and was previously thought to be HIV-1_{BRU}. (For an explanation of the misassignment and summary of the history of the two strains, refer to Wain-Hobson *et al.*, 1991). HIV-1_{MAL} was isolated in 1985 from a 7-year-old Zairean male with ARC (AIDS Related Complex) who is thought to have acquired the virus through blood transfusion since both parents were seronegative. HIV-1_{ELI} was obtained in 1983 from a 24-year-old Zairean female with AIDS. The sequences of these three viruses have been determined: HIV-1_{LAI} by Wain-Hobson *et al.* (1985) and HIV-1_{MAL} and HIV-1_{ELI} by Alizon *et al.* (1986). However, because complete proviral genomes were not cloned, in no case was the sequence shown to encode infectious virus.

For our continuing molecular genetic analyses of HIV-1, it was necessary to have infectious molecular clones of viruses whose sequences were known. In addition, a comparison of the biological activities of viruses obtained from different geographical areas, from patients at various stages of HIV disease, and whose sequences are divergent might reveal differences that can be correlated with biological activity and perhaps pathogenesis. Therefore, we have assembled full-length copies not only of the original HIV-1 isolate, LAI, but also of two central African viruses, HIV-1_{MAL} and HIV-1_{ELI}. We report here the initial characterization of the viruses derived from these molecular clones. It was found that, although all three viruses were infectious in peripheral blood mononuclear cells (PBMC), their growth kinetics differed and the three exhibited different tropisms in established CD4$^+$ cell lines. Furthermore, passage of HIV-1_{LAI} and HIV-1_{ELI} in

[1] To whom reprint requests should be addressed at Laboratory of Molecular Microbiology, Building 4, Room 310, NIAID, NIH, Bethesda, MD 20892.

[2] Present address: Fred Hutchinson Cancer Research Center, Division of Basic Sciences, 1124 Columbia Street, Seattle, WA 98104.

0042-6822/91 $3.00
Copyright © 1991 by Academic Press, Inc.
All rights of reproduction in any form reserved.

culture was found to result in the appearance of viruses that had altered growth potential, while passage of HIV-1_{MAL} did not change its properties. The viruses that emerged from the growth of ELI on CEM and H9 cells had accelerated growth kinetics on those two cell lines and in addition had a broadened host range. On the other hand, passage of LAI through CEM and U937 cells diminished its replication capacity on SupT1 cells. These results suggest that the high degree of sequence variability of HIV is not due solely to immune selection, but may also be due to selection for growth in different $CD4^+$ cells.

MATERIALS AND METHODS

Bacterial strain, plasmids, and phage DNA

Throughout this work, the *Escherichia coli* strain DH5 was used. It was purchased as competent cells from BRL (Gaithersburg, MD) and transformed according to Hanahan (1983). The HIV-1_{LAI} clone λJ19 was described in Alizon *et al.* (1984) and Wain-Hobson *et al.* (1985) and was isolated as a partial *Hin*dIII fragment from the DNA of $CD4^+$ lymphocytes infected with the LAI (LAV) isolate of HIV-1. The plasmid pLTR, which contains a complete copy of the LTR from nt 8068 (*Bam*HI) to 505 (*Acc*I), was constructed from λJ19 by Marc Alizon. Two deletion mutant derivatives of pBR322 were used in this work: pKP59 retains nt 2457 to the *Eco*RI site at 4361 and has a polylinker with sites for *Eco*RI, *Avr*II, *Pst*I, *Pvu*II, *Kpn*I, *Bam*HI, *Cla*I, *Hin*dIII, *Nde*I, *Sac*I, *Xho*I, *Sal*I, *Bgl*II, *Mlu*I, *Hpa*I, and *Xba*I (Mounts *et al.*, 1989) and pKP55 has a polylinker with sites for the enzymes *Cla*I, *Hin*dIII, *Nco*I, *Sph*I, *Pvu*II, *Kpn*I, and *Xba*I (Peden, unpublished).

Construction of full-length proviral clones

For the assembly of the proviral genomes, the related plasmids pKP59 and pKP55 were used as the vectors. Due to the number of steps involved, details are not presented but can be obtained from K.P. upon request. In all cases, no DNA other than that originally sequenced was used, and all of the reading frames were maintained. Since at least two clones were assembled for each strain and viruses derived from replicate clones were found to have similar properties, it is unlikely that mutations had been introduced during cloning.

The lambda clone J19 was the sole source of HIV-1_{LAI} DNA used in the construction of pLAI3, the full-length genomic copy of this virus. The plasmid pLTR was the source of both LTR elements and was constructed by Marc Alizon, Institut Pasteur, from λJ19 by ligating the 5′ and 3′ parts at the *Hin*dIII site. In a series of steps, pLAI3 was assembled from pLTR and λJ19, and its structure is shown in Fig. 1. At the 5′ end there are 34 additional nt from the *nef* gene from the *Bgl*II site at nt 8644 before the U3 sequences, and the 3′ end is at the *Nar*I site at nt 182 in the primer binding site (PBS). (Since the *Nar*I site in LAI was ligated to the *Cla*I site in the vector, both sites were lost.) Therefore, pLAI3 consists of the 9766 bp of the proviral genome, 34 bp due to the duplication at the 5′ end, and the 2 kb of vector resulting in a plasmid of approximately 11.8 kb in size. The Pol^- version of pLAI3, pLAI3pol1, was prepared by deleting the *Pvu*II fragment (nt 3335 to 4379) in the *pol* gene.

The lambda clone M-H11 (Alizon *et al.*, 1986) was the starting material for pMAL2. From this clone, a permuted derivative, pME200, was constructed by digesting with *Eco*RI and *Hin*dIII, ligating at the *Hin*dIII site, and cloning into pKS(+) (Stratagene, La Jolla, CA) at its *Eco*RI site. The Asp718 to *Nar*I fragment (nt 9063 to 636) was cloned into pKP55 between the Asp718 and *Cla*I sites to produce $pLTR_{MAL}$. Ligation of the *Hin*dIII fragment from pME200 into $pLTR_{MAL}$ resulted in pMAL2 (Fig. 1). This plasmid consists of 9767 bp of HIV-1_{MAL}, 72 bp from the *nef* gene at the 5′ end in a vector of 2 kb resulting in a size of 11.8 kb.

For the assembly of pELI1, the lambda clone E-H12 (Alizon *et al.*, 1986) was the source of the DNA. This phage clone contains a complete 5′ LTR and part of the 3′ LTR truncated at the *Hin*dIII site; λE-H12 was cloned after partial *Hin*dIII digestion of infected cell DNA. The 5′ LTR of E-H12 was used for both 5′ and 3′ LTRs. In a series of steps, the LTR elements were duplicated and viral sequences inserted between them to produce pELI1, a plasmid consisting of about 1.3 kb of cellular DNA upstream of the 5′ LTR and 9713 bp of HIV sequences in a vector of 2 kb resulting in a plasmid of 12.8 kb (Fig. 1). As was the case for the other two genomes, the 3′ end of ELI is at the *Nar*I site.

The sequence of pELI from nt 7362 to 7920 was determined using the chain termination method (Sanger *et al.*, 1977) adapted for use with plasmid DNA (Chen and Seeburg, 1985) and with [α-^{35}S]dATP as the labeled nucleotide (Biggin *et al.*, 1983). Sequenase version 2.0 (United States Biochemical Corp., Cleveland, OH) was used as the DNA polymerase.

Preparation of plasmid DNA

Plasmid DNA from both small scale (1.5 ml) and large scale (300 ml) cultures was prepared using the lysis procedure of Ish-Horowicz and Burke (1981). For large scale preparations, the DNA was purified by twice banding in CsCl–ethidium bromide gradients,

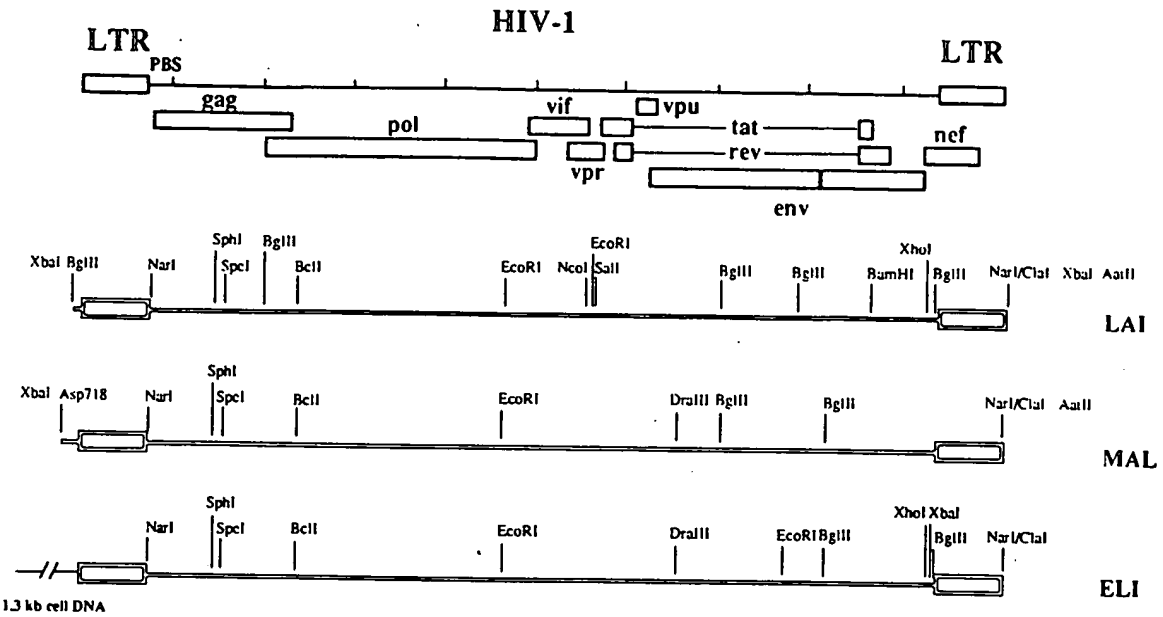


FIG. 1. Structures of the genomes of LAI, MAL, and ELI. The top of the figure shows the approximate positions of the known open reading frames of HIV-1 and the PBS. Beneath the viral and cellular DNA sequences of the three proviral clones are shown; vector sequences have been omitted. Thick lines are used for the viral sequences and thin ones for cellular DNA. Only the locations of some of the restriction enzyme sites are indicated.

proteinase K digestion, phenol extractions, and alcohol precipitations.

Cells and media

HeLa cells were obtained from Barrie Carter (NIH) and were grown in Dulbecco's modified Eagle's medium with 10% fetal bovine serum (FBS). CD4-positive lymphocyte cell lines SupT1, the CEM derivative clone 12D7, HUT-78 and its derivative H9, and the promonocyte cell line U937 were grown in RPMI 1640 with 10% FBS. SupT1 cells (Smith *et al.*, 1984) were obtained from the AIDS Research and Reference Program of the NIH; CEM clone 12D7 was derived by G. Poli (NIH) from CEM cells (Foley *et al.*, 1965); Jurkat clone E6-1 (Weiss *et al.*, 1984) was obtained from ATCC, and the HUT-78 (Gazder *et al.*, 1980) derivative, H9 (Mann *et al.*, 1989), was obtained from the FDA; U937 cells (Sundstrom and Nilsson, 1976) were obtained from the Pasteur Institute. PBMC were obtained either from Microbiological Associates (Gaithersburg, MD) or from SRA Technologies (Alexandria, VA) and had been prepared from HIV-1 seronegative, healthy donors and purified by centrifugation through a Ficoll–Hypaque cushion. Before use, they were stimulated with phytohemaglutinin (PHA) (0.25 μg/ml) for 72 to 96 hr and grown in RPMI 1640 with 10% FBS and 10% IL-2 (Pharmacia Diagnostics, Fairfield, NJ).

Assays for biological activity

HeLa cells (5×10^5 cells) in T25 flasks were transfected with plasmid DNA (15 μg) using calcium phosphate coprecipitation methods (Graham and van der Eb, 1973; Frost and Williams, 1978; Wigler *et al.*, 1979). Reverse transcriptase activity in the culture medium was used to monitor virus production; when it was tested, p24 determinations gave comparable results. Reverse transcriptase activity (Willey *et al.*, 1988) was measured 24, 48, and 72 hr after transfection. Supernatants generally contained maximum activity between 24 and 48 hr and then decreased. For the infection of cell lines, various amounts of virus corresponding to a volume of supernatant that resulted in the production of $2–5 \times 10^5$ cpm of [^{32}P]TTP incorporated by the viral reverse transcriptase using oligo(dT) · poly(A) as template in a standard assay were added to the cells in RPMI 1640 plus 10% FBS and 2 μg/ml of polybrene. Virus adsorbtion was for 2 hr at 37°, after which time the cells were diluted in RPMI 1640 with 10% FBS and 2 μg/ml polybrene (RPMI 10/2). At 2-day intervals, the cells were fed with RPMI 10/2 and samples taken for reverse transcriptase activity determination.

RESULTS

Biological activity of infectious molecular clones of LAI, MAL, and ELI on primary cells

Although the original clones for LAI, MAL, and ELI contained all the genetic information for these viruses (Wain-Hobson *et al.*, 1985; Alizon *et al.*, 1986), none of the clones contained two complete copies of the LTR, and therefore their infectivities could not be readily assessed or compared. Through a series of steps, proviral copies of all three viruses containing two LTR elements were assembled (see Materials and Methods).

To measure the infectivity of LAI, MAL, and ELI, monolayer cultures of HeLa cells were transfected with 15–20 μg of the respective plasmid DNA, and at 24, 48, and 72 hr culture medium was withdrawn and the amount of virus determined by assaying for virion-associated reverse transcriptase activity. Maximum production was usually at 48 hr after transfection. Based on the reverse transcriptase activity, equal amounts of virus were used to infect PBMC that had been stimulated with PHA (0.25 μg/ml) for 3–4 days. The cultures were monitored for the appearance of syncytia as well as for the production of virus, as indicated by reverse transcriptase activity released into the culture medium.

Syncytia were apparent in the cultures infected with LAI and ELI on Day 3 after infection. For LAI, their number and size increased until Day 5 followed by a decrease due to lysis. The kinetics of syncytium development and decay were similar for ELI, although the syncytia were larger. No syncytia were seen with MAL. Production of virus for LAI and ELI closely paralleled syncytium formation, reaching a maximum on Day 6 and declining thereafter (Fig. 2). For MAL, the peak of reverse transcriptase activity was later, between Days 10 and 12, and overall less virus was produced (Fig. 3).

The published sequence of ELI appeared anomalous in comparison with other HIV-1 isolates (Myers *et al.*, 1991) in the region previously shown to be important for CD4 binding (Laskey *et al.*, 1987). In particular, all HIV-1 envelopes have the sequence Asn–X–Trp–Gln (residues 430 to 433 relative to the Env of HIV-1_{LAI}), where X is usually Met. It has been shown that the Trp at 432 is essential for the binding of Env to CD4 and for viral infectivity (Cordonnier *et al.*, 1989). Because the virus derived from pELI1 was infectious in PBMC, the sequence of this region of the *env* gene was determined. It was found that the sequence differred from the published sequence as follows: G at 7038 should be C; G at 7041 is not present; there is an insertion of GGA between 7048 and 7049; and an insertion of A between 7053 and 7054. These changes do not alter the reading frame but modify the sequence of the *env* gene and its predicted protein. The region of the *env* gene from nt 7487 to 7519 and the predicted Env proteins of ELI, LAI, and MAL from residues 422 to 433 are shown in Fig. 3. With these corrections, the Env of ELI now resembles those of other HIV-1 isolates (see Myers *et al.*, 1991).

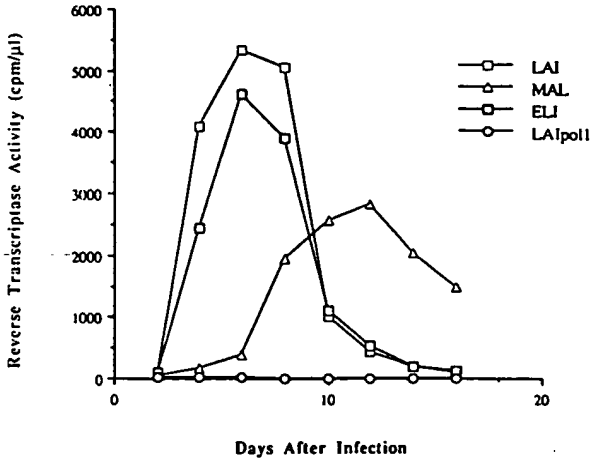


FIG. 2. Growth of LAI, MAL, and ELI on peripheral blood mononuclear cells. Approximately 2 × 10^6 PHA-stimulated PBMC were infected with supernatants corresponding to 2 × 10^5 reverse transcriptase cpm as described under Materials and Methods. A similar amount of the transfection supernatant from pLAIpol1 was used as a negative control. Every day the cultures were monitored for the appearance of syncytia and every 2 days the cultures were fed and samples were taken for reverse transcriptase activity determinations.

Biological activity on established T cell lines

Four CD4-positive T cell lines were used—the 12D7 clone of the CEM line, the H9 clone of the HUT-78 line, the SupT1 line, and the E6-1 clone of Jurkat cells. Supernatants from transfected HeLa cells were prepared as described in the preceding section, and equal amounts of virus according to the number of reverse transcriptase counts were used for the infections. The infections were monitored both for the appearance of syncytia in the culture and for the production of virus in the medium.

The three viruses displayed different tropisms for the T cell lines (Fig. 4). CEM cells were readily permissive for LAI with reverse transcriptase activity appearing in the culture supernatant on Day 4 after infection, reaching peak production around Day 8, and diminishing thereafter (Fig. 4A). The time course of syncytium formation in the culture and subsequent cell lysis corresponded to the release of reverse transcriptase activity. In the case of ELI, however, no virus production was seen on CEM cells until about 30 days after infection, and there was little apparent syncytium formation.

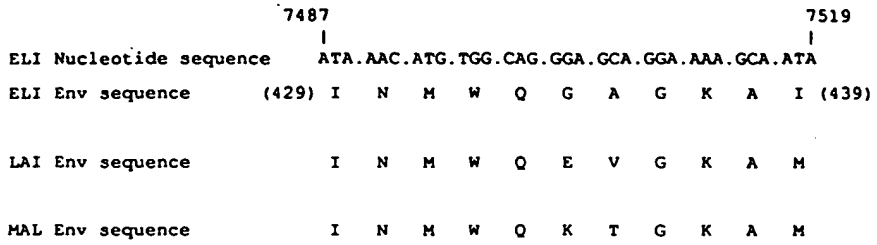


FIG. 3. DNA sequence of ELI in the CD4 binding region together with the predicted protein sequences of ELI, LAI, and MAL. The nucleotide sequence from 7487 to 7519 is shown at the top. Beneath are the predicted protein sequences for ELI, LAI, and MAL from residues 429 to 439 (numbers are relative to the LAI sequence).

With MAL, no virus production could be detected in the cultures even after 55 days (Fig. 4A).

H9 cells were readily infectable with LAI, where reverse transcriptase activity was first detected on Day 4 after infection, concomitant with the appearance of large syncytia in the culture (Fig. 4B). Peak virus production was on Day 6 and, after falling until Day 20, began to rise again until Day 31 before falling once again. This periodicity in virus production has been frequently seen after establishment of a chronic infection in H9 cells. The kinetics of ELI infection showed that virus production was delayed compared to that of LAI, with reverse transcriptase activity first detectable on Day 23 and rising to a maximum around Day 30. As was the case with LAI, a chronic infection appeared to have been established. MAL was not infectious in H9 cells (Fig. 4B).

MAL and LAI were infectious in SupT1 cells, but ELI was not (Fig. 4C). Both the kinetics of virus production and the amount of virus produced for LAI and MAL were similar, with both viruses reaching peak production between Days 18 and 21. Virus production was often lower in SupT1 cells than in the other lines tested. In fact, the kinetics of infection of LAI on SupT1 cells were frequently variable from experiment to experiment, unlike that seen on other cells. Large and abundant syncytia were present in the LAI-infected cultures but no syncytia were seen with MAL.

LAI and ELI were infectious in Jurkat cells and both viruses induced the formation of large syncytia. The time course of LAI infection preceded that of ELI by several days. This cell line was not permissive for MAL.

An additional T cell line was used, the HTLV-1 transformed line MT-4. LAI and ELI replicated on this line but MAL could not (data not shown).

Biological activity on an established promonocyte cell line

The three viruses had different growth properties on the U937 human promonocyte cell line. Only LAI was able to infect these cells, with reverse transcriptase activity becoming apparent in the culture on Day 6 following infection and reaching a maximum around Day 15 (Fig. 5). No discernible syncytia were ever observed. Neither MAL nor ELI could be grown on U937 cells. In the experiment shown in Fig. 5, a chronic infection was established for LAI, although this does not always occur. The capacity to form a chronically HIV-infected culture of U937 is more variable than with H9 cells, where almost all infections with the LAI isolate of HIV-1 have this outcome.

The biological activities of LAI, MAL, and ELI on the different cells are summarized in Table 1.

Passage of LAI on CD4-positive T cell lines

To determine whether the phenotypes of the viruses were stable to passage on different cell lines, virus stocks prepared on one particular cell line were used to infect various T cell lines and the promonocytic U937 cell line. LAI was propagated on the CEM, SupT1, Jurkat, and U937 cell lines and equal amounts of virus were used to infect those cell lines and also H9 cells. After passage of LAI through each of the cell lines, there were significant alterations to the growth kinetics and virus production (Fig. 6). Most notably, when virus stocks prepared on CEM, SupT1, Jurkat, and U937 cells were used to infect SupT1 cells, there was a marked delay in the kinetics of viral growth with the CEM and U937 stocks compared to the virus stocks prepared on SupT1 and Jurkat cells (Fig. 6C). Also, virus prepared on Jurkat cells replicated significantly better on these cells than viruses propagated on CEM, SupT1, or U937 cells (Fig. 6D). Finally, LAI passaged through SupT1 cells consistently produced less virus on CEM, H9, Jurkat, and U937 cells than did virus prepared on the other cells.

These results demonstrate that the biological activity of LAI is modified by passage through different cell lines.

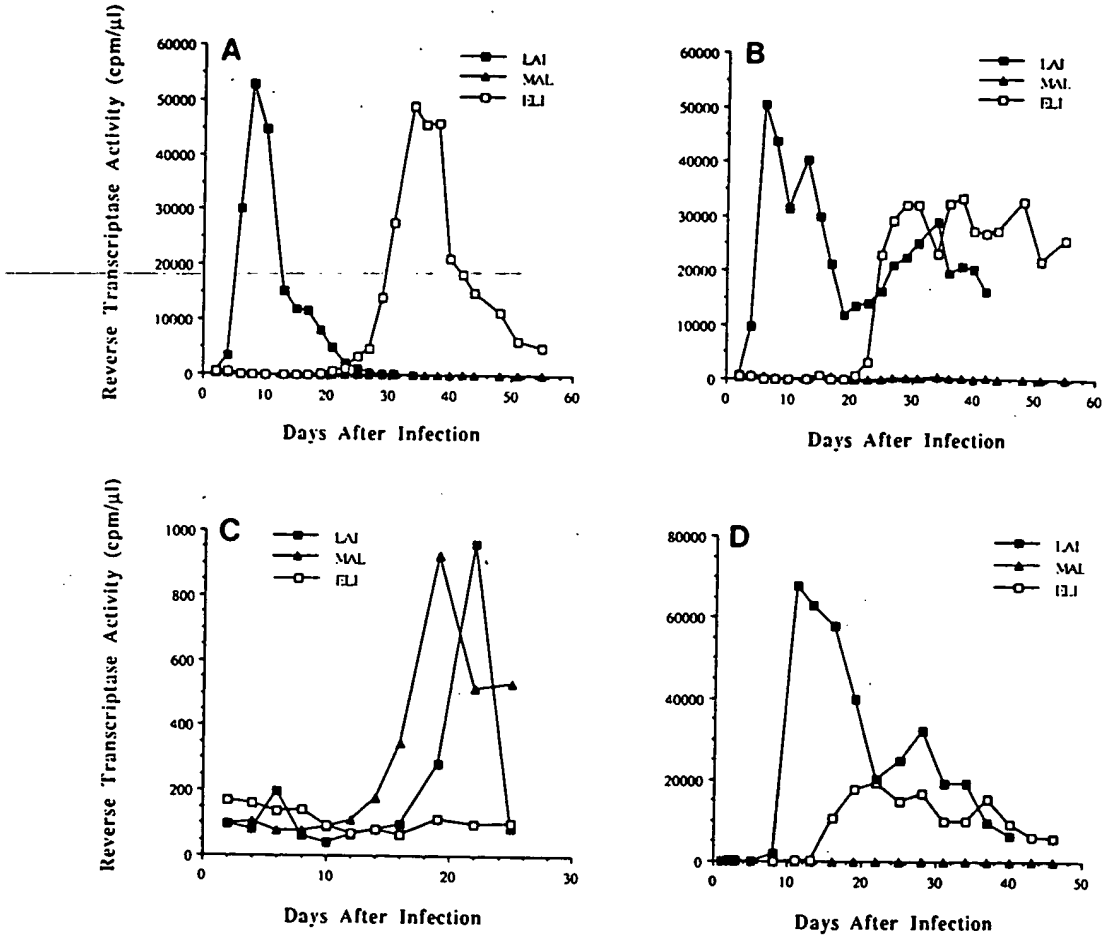


FIG. 4. Growth of LAI, MAL, and ELI on T cell lines. Approximately 2 × 10^6 CEM, H9, SupT1, and Jurkat cells were infected with supernatants corresponding to 2 × 10^5 reverse transcriptase cpm as described under Materials and Methods. Every day the cultures were monitored for the appearance of syncytia and every 2 days the cultures were fed and samples were taken for reverse transcriptase activity determinations. (A) Infection of CEM cells; (B) infection of H9 cells; (C) infection of SupT1 cells; (D) infection of Jurkat cells. The infection of SupT1 cells in this particular experiment produced low virus yields. Infection of this cell line produces variable amounts of virus. In the experiments shown in Figs. 6 and 7, greater amounts of virus were produced.

Serial passage of ELI *in vitro* results in a virus with improved growth potential

To ascertain whether the virus that appeared in the cultures of ELI-infected CEM and H9 cells after 25 days (Figs. 4A and 4B) exhibited similar delayed growth kinetics when used to reinfect these cells, virus was isolated from the H9 cultures (ELI.H9) and used to infect CEM, H9, SupT1, and U937 cells. LAI prepared on H9 cells (LAI.H9) was used as a reference for the infection kinetics. Comparing the replication of ELI.H9 to that of LAI.H9 was considered more valid than comparing ELI.H9 to ELI prepared after transfection of HeLa cells, since the cells used to prepare the virus stocks can influence the phenotype of the resulting virus. LAI was chosen as the reference since neither the growth kinetics nor the tropism of LAI was significantly affected by passage through H9 cells. For example, the peak virus production assayed by reverse transcriptase activity was around Day 10 in both cases on CEM cells (Figs. 4A and 7A), around Day 8 on H9 cells (Figs. 4B and 7B), about Day 20 on SupT1 cells (Figs. 4C and 7C), and between Days 15 and 20 on U937 cells (Figs. 5D and 7D). In contrast, ELI passaged through H9 cells (ELI.H9) now had accelerated growth kinetics on CEM and H9 cells compared to ELI prepared from transfection of HeLa cells; peak virus production was advanced from between Days 30 and 35 (Fig. 4A) to between Days 15 and 18 (Fig. 7A) on CEM cells and from Days 25 to 30 (Fig. 4B) to Day 9 (Fig. 7B) on H9 cells. This

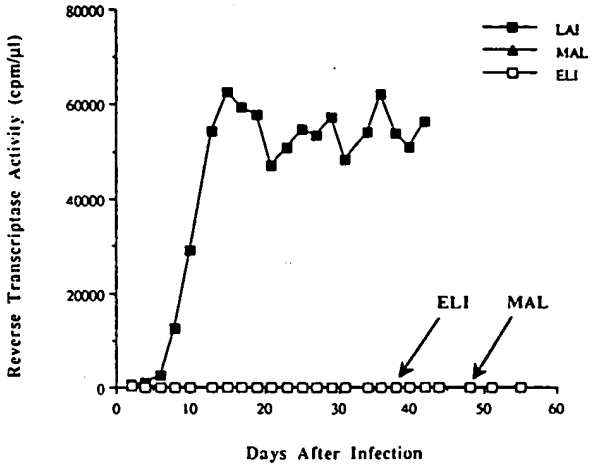


FIG. 5. Growth of LAI, MAL, and ELI on the U937 promonocytic cell line. Approximately 2×10^6 U937 cells were infected with supernatants corresponding to 2×10^5 reverse transcriptase cpm as described under Materials and Methods. Every day the cultures were monitored for the appearance of syncytia and every 2 days the cultures were fed and samples were taken for reverse transcriptase activity determinations.

enhanced replicative capacity of ELI.H9 was found even if different dilutions of the virus stocks were used. In addition, ELI.H9 was now able to grow on U937 cells, which were previously refractory to infection by ELI (compare Fig. 5 and Fig. 7D), and after more than 50 days in culture was even able to infect SupT1 cells (data not shown). The properties of LAI and ELI resulting from the passage through H9 cells are summarized in Table 1.

The second peak of virus production seen after infection of H9 cells (Fig. 7B) is due to the establishment of a chronic infection in these cultures. In the case of ELI, it is not due to the replication of the original virus present in the transfection stock, as virus obtained from both peaks was found to have similar properties when tested on a variety of cell lines.

In contrast to that seen with LAI and ELI, the phenotype of MAL was not changed by passage through SupT1 cells, since MAL and MAL.SupT1 do not have significantly different growth kinetics or extended host range on CEM, H9, SupT1, and U937 cells (data not shown).

DISCUSSION

Although the complete genomic sequences for several HIV-1, HIV-2, and SIV isolates have been determined (for a compilation and analysis see Meyers *et al.*, 1991), not all of the molecular clones for which a sequence is known have been shown to code for infectious viruses. In most cases, this was because a complete proviral genome containing two LTR elements was not obtained in the original clone or the genome was isolated in two or more fragments. For others, termination codons in essential genes were revealed by nucleotide sequencing, and these most likely caused the defect in those clones. For other clones, the nature of the defect was not obvious and may be a result of the use of inappropriate cells to propagate the virus.

To ascertain whether the published sequences of LAI, MAL, and ELI coded for infectious viruses, two LTR proviral genomes were assembled for each isolate. One reason for putting emphasis on obtaining several infectious molecular clones whose sequences are known is that it is hoped that biological activities such as cytopathogenicity, replication capacity, and cell tropism can be correlated with genomic sequence differences among the isolates, and this may implicate genes and/or regulatory regions involved in those activities. Importantly, these three viral genomes possess all known open reading frames, although the MAL clone has an isoleucine codon in place of the initiator

TABLE 1

BIOLOGICAL ACTIVITIES OF HIV ISOLATES IN DIFFERENT CELL TYPES

HIV-1 isolate	Cell type						
	PBMC	CEM	H9	SupT1	Jurkat	MT-4	U937
LAI	+++,s	+++,s	+++,s	++,s	+++,s	+++	+++
MAL	++	–	–	++	–	–	–
ELI	+++,s	d,+++,s	d,+++,s	–	+++,s	+++	–
LAI.H9	+++,s	+++,s	+++,s	++,s	+++,s	+++	+++
ELI.H9	+++,s	+++,s	+++,s	d,+++,s	+++,s	+++	++

Note. +, virus growth detectable by reverse transcriptase activity; –, no reverse transcriptase activity detected; d, delayed appearance of virus; s, presence of syncytia in cultures.

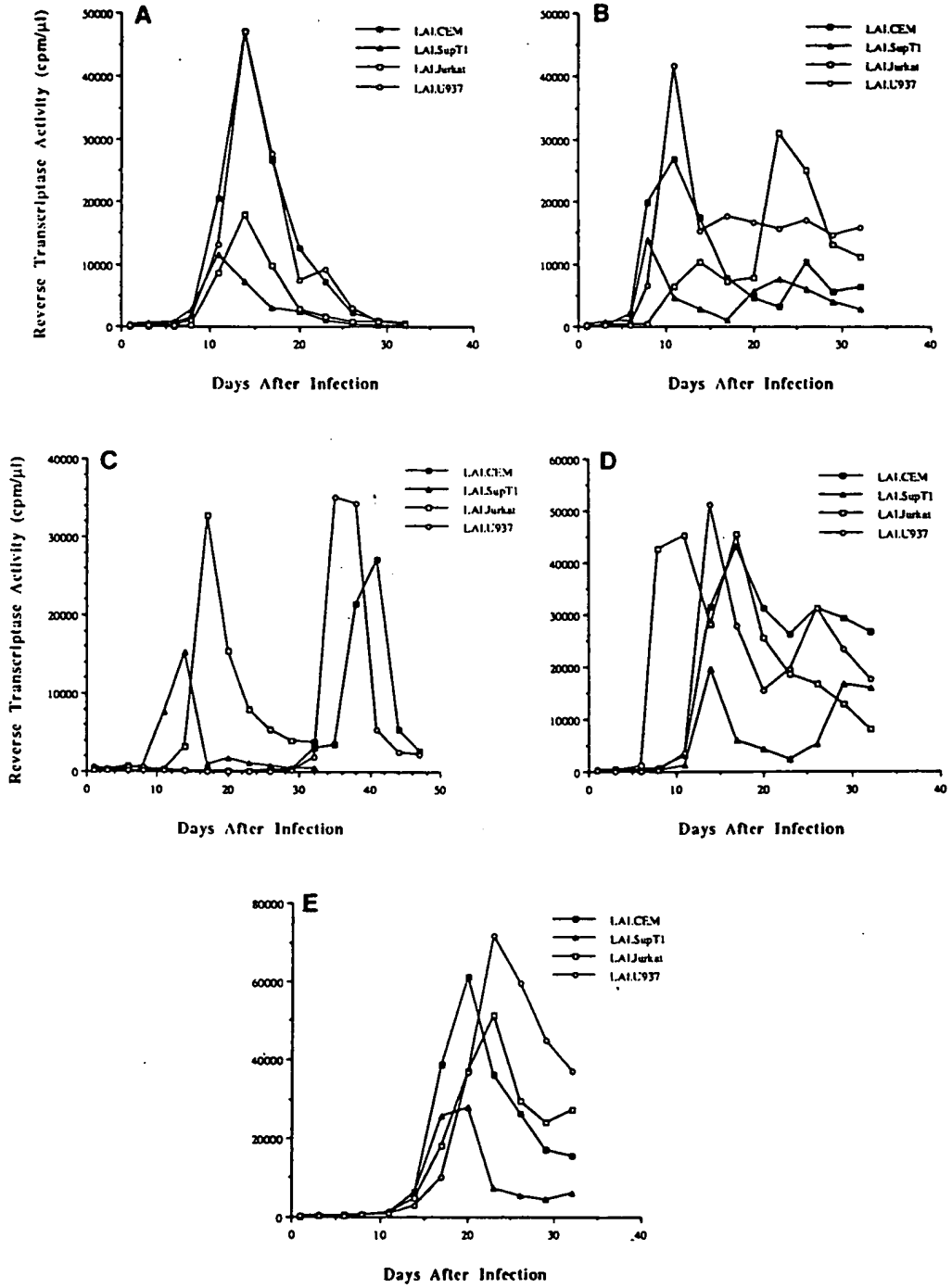


FIG. 6. Kinetics of virus replication in CEM, H9, SupT1, Jurkat, and U937 cells of LAI passaged in CEM, SupT1, Jurkat, and U937 cells. Supernatants corresponding to 5×10^5 reverse transcriptase cpm used to infect 2×10^6 cells of various lines: (A) CEM cells infected with LAI.CEM, LAI.SupT1, LAI.Jurkat, and LAI.U937. (B) H9 cells infected with LAI.CEM, LAI.SupT1, LAI.Jurkat, and LAI.U937. (C) SupT1 cells infected with LAI.CEM, LAI.SupT1, LAI.Jurkat, and LAI.U937. (D) Jurkat cells infected with LAI.CEM, LAI.SupT1, LAI.Jurkat, and LAI.U937. (E) U937 cells infected with LAI.CEM, LAI.SupT1, LAI.Jurkat, and LAI.U937.

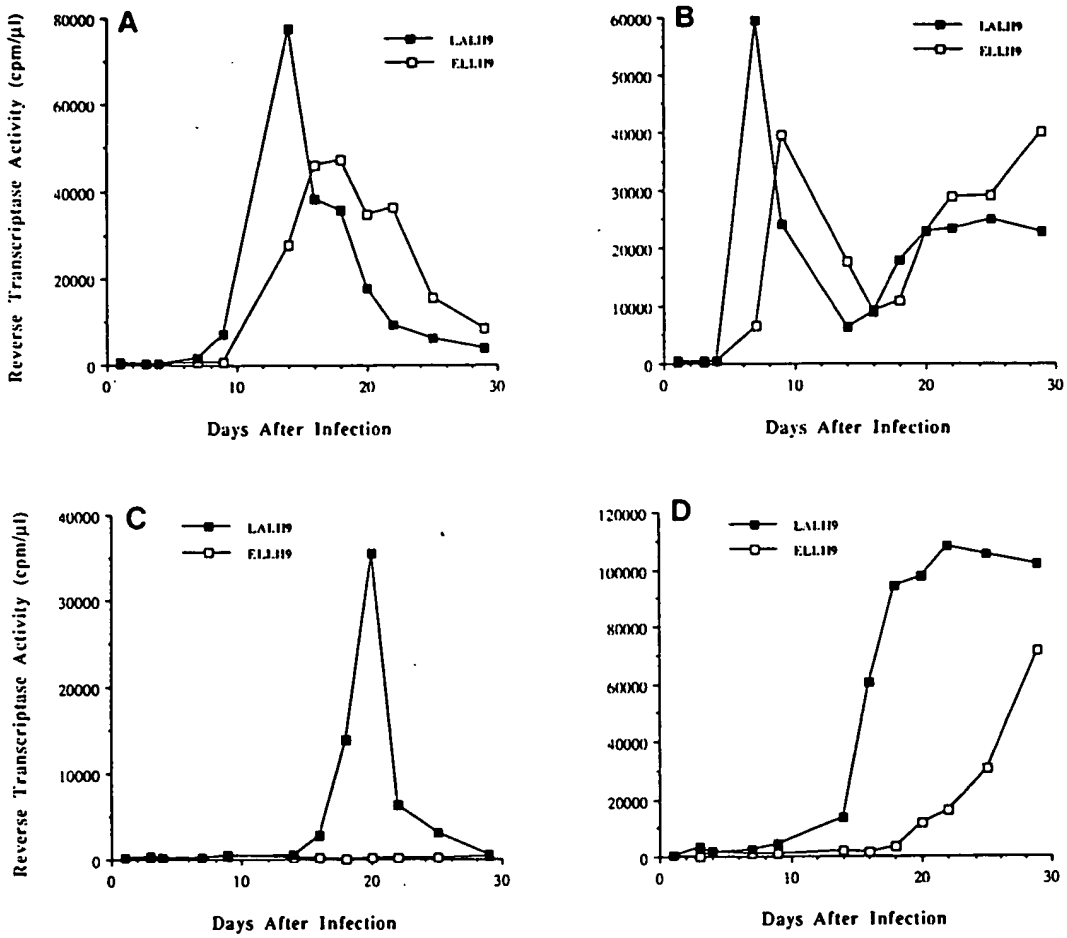


FIG. 7. Kinetics of infection on CEM, H9, and U937 cells of ELI and LAI passaged in H9 cells. Virus stocks of LAI and ELI were prepared from H9 cells and supernatants corresponding to 5×10^5 reverse transcriptase cpm used to infect various cell lines: (A) infection of CEM cells with LAI.H9 and ELI.H9; (B) infection of H9 cells with LAI.H9 and ELI.H9; (C) infection of SupT1 cells with LAI.H9 and ELI.H9; (D) infection of U937 cells with LAI.H9 and ELI.H9.

methionine codon and may thus be Vpu$^-$ or at least produce less Vpu. The integrity of all open reading frames is not found with some clones. For example, the commonly used infectious HxB2 clone has threonine as the initiator codon for Vpu, and the Vpr and Nef proteins are truncated due to frameshifts in the *vpr* and *nef* genes, respectively. Therefore, unless these three genes are corrected, the interpretation of the results obtained on the analysis of viruses with mutations in other genes must be considered in this light, namely, that of multiple mutants.

Viruses produced after transfection of HeLa cells were assessed for their infectivity on PBMC and a variety of T cell lines and the promonocyte cell line, U937. On both primary peripheral blood mononuclear cells and CD4-positive cell lines of both the lymphocytic (CEM, H9, SupT1, Jurkat, MT-4) and monocytic (U937) lineages, LAI produces a rapid spreading infection. On PBMC, CEM, H9, SupT1, and Jurkat cells, LAI induces the formation of syncytia in the cultures and can therefore be considered to be of the rapid/high and syncytium inducing (SI) class (Fenyö *et al.*, 1988; Tersmette *et al.*, 1988). ELI should also be classified as rapid/high, since its replicative capacity on PBMC is high, it replicates in Jurkat and MT-4 cells, and in H9 and CEM cells although is delayed relative to LAI, and also induces syncytia in the cultures. This is consistent with the fact that both LAI and ELI were obtained from people with AIDS, since a correlation has been found between viruses isolated at different stages of HIV disease and their properties *in vitro*. For example, viruses isolated from infected people in the asymptomatic

phase are more likely to be less cytopathic and nonsyncytium inducing (NSI) (Cheng-Mayer *et al.*, 1988; Tersmette *et al.*, 1988; Fenyö *et al.*, 1988; Åsjö *et al.*, 1990) and have restricted capacity to infect T cell lines —the slow/low, NSI class (Tersmette *et al.*, 1988; Fenyö *et al.*, 1988). In fact, many such isolates are only able to replicate on primary cells and are unable to infect cell lines (Schuitemaker *et al.*, 1991; Groenink *et al.*, 1991). At the later stages of disease, however, virus isolates are more likely to have an expanded host cell range and be able to infect cell lines, to be more cytopathic and induce the production of syncytia, and to replicate to higher levels—the rapid/high type (Cheng-Mayer *et al.*, 1988; Tersmette *et al.*, 1988; Fenyö *et al.*, 1988; Åsjö *et al.*, 1990).

MAL, however, is probably a member of the slow/low, NSI class of the group 3 type (Fenyö *et al.*, 1988; Åsjö *et al.*, 1990), which replicates in PBMC and in some cell lines but does not induce syncytia. This virus was isolated from a patient with ARC.

The observation that ELI passaged on different cell lines had increased replication capacity and extended host range was unexpected. Two mechanisms could account for these phenotypic alterations. First, the HIV genome could be less stable during propagation *in vitro* than hitherto realized, and the mutation rate is high enough in tissue culture to allow for the selection of variants better adapted for growth on a particular cell line. Such a high rate of mutation presumably accounted for the appearance of revertants of defective HIV mutants with changes introduced into the envelope gene (Willey *et al.*, 1988, 1989). Second, epigenetic mechanisms could be operating, such as to produce a virion with increased replication ability, possibly through modification of virion proteins or the RNA genome or the presence of host cell components in the viral envelope. Similar observations on the adaptation of viruses to various cells on which they were grown have been made recently by Cheng-Mayer *et al.* (1991) and Fredriksson *et al.* (1991). Recent results from Vartanian *et al.* (1991) favor the hypothesis that the virus has undergone genetic changes, since they demonstrated that a single passage on two cell lines of virus isolated from a patient resulted in the generation of many genetically distinct variants. Furthermore, for an epigenetic mechanism to account for the increased replication potential of the virus, it would have to be operating on the initial round of infection, since once the virus had undergone one round of growth, the virion would acquire the modification of that host cell and no restriction should exist for subsequent infectious cycles. The kinetics of infection seem to argue against this type of mechanism.

Another, less likely, type of mechanism may involve the presence of cell type-specific cytokines in the viral stock that promote cellular and/or viral growth. For this type of mechanism to operate after the initial infectious cycle, one would have to postulate that the cytokine could induce its own synthesis and set up an autocrine loop.

The sudden and rapid appearance of ELI after a 30-day lag is consistent with the acquisition of genetic changes that permit virus growth on those cells. For LAI, mutation and subsequent selection could also be occurring in the case where viral stocks prepared on CEM or H9 cells exhibit delayed growth kinetics on SupT1 cells.

Thus, we have described the assembly and properties of three additional infectious molecular clones of HIV-1 prepared from viruses isolated originally from central Africa (MAL and ELI) and North America (LAI). Viruses derived from these clones have different replicative abilities on PBMC and tumor cell lines. It will be informative to elucidate the molecular determinants that account for these differences in tropism and to ascertain the step at which viral replication is blocked in the nonpermissive cells. Since these cell lines were derived from T cell (CEM, H9, SupT1, and Jurkat) or a histiocytic tumor (U937), they presumably represent different stages in the ontogeny of lymphocytes and monocytes, and thus determining the mechanism of the cell tropism may be informative as to the types of cells in the developmental pathways of CD4-positive cells that can be infected *in vivo*. It will be of interest, therefore, to test the tropism of these three viruses on different subsets of untransformed CD4-positive primary T cells as well as on monocyte/macrophages. In this paper, we have demonstrated that the phenotypes of HIV-1 *in vitro* are less stable than previously realized and that growth of some HIVs derived from infectious molecular clones on various cell lines can result in viruses with increased growth potential and expanded host range. While the mechanism for such adaptation is unknown, it is presumably due to subtle selection pressures imparted by the particular cells on viral replication. This indicates that the sequence variability of HIV strains is not due to immune selection alone. Rather, selection for growth in particular cells is enough to fix genetic changes in the absence of an immune response. The mechanism of such tissue culture adaptation may be relevant to the development of more pathogenic viruses during disease progression *in vivo* and may provide an *in vitro* model in which to study the events leading to the appearance of more virulent variants in the absence of immune selection of the host.

ACKNOWLEDGMENTS

We thank Phoebe Mounts, Malcolm Martin, Ron Willey, Kuan-Teh Jeang, Hironori Sato, and M. David Hoggan for comments on the manuscript and Malcolm Martin for support. K.P. expresses gratitude to the U.S. INS for their cooperation in allowing these studies to be carried out. This work was supported by Contract MD800803 from the NIH, by a fellowship from the Leukemia Society of America to M.E., and by a fellowship from la Fondation pour la Recherche Médicale to K.P.

REFERENCES

ÅSJÖ, B., SHARMA, U. K., MORFELDT-MÅNSON, L., MAGNUSSON, A., BARKHEM, T., ALBERT, J., OLAUSSON, E., VON GEGERFELT, A., LIND, B., BIBERFELD, P., and FENYÖ, E. M. (1990). Naturally occurring HIV-1 isolates with differences in replicative capacity are distinguished by *in situ* hybridization of infected cells. *AIDS Res. Human Retroviruses* 6, 1177–1182.

ALIZON, M., SONIGO, P., BARRÉ-SINOUSSI, F., CHERMANN, J.-C., TIOLLAIS, P., MONTAGNIER, L., and WAIN-HOBSON, S. (1984). Molecular cloning of lymphadenopathy-associated virus. *Nature* 312, 757–760.

ALIZON, M., WAIN-HOBSON, S., MONTAGNIER, L., and SONIGO, P. (1986). Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients. *Cell* 46, 63–74.

BARRÉ-SINOUSSI, F., CHERMANN, J.-C., REY, F., NUGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C., BRUN-VÉZINET, F., ROUZIOUX, C., ROZENBAUM, W., and MONTAGNIER, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome. *Science* 220, 868–870.

BIGGIN, M. D., GIBSON, T. J.,20and HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80, 3963–3965.

CHEN, E. Y., and SEEBURG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* 4, 165–170.

CHENG-MAYER, C., SETO, D., TATENO, M., and LEVY, J. A. (1988). Biologic features of HIV-1 that correlate with virulence in the host. *Science* 240, 80–82.

CHENG-MAYER, C., SETO, D., and LEVY, J. A. (1991). Altered host range of HIV-1 after passage through various human cell types. *Virology* 181, 288–294.

CORDONNIER, A., MONTAGNIER, L., and EMERMAN, M. (1989). Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. *Nature* 340, 571–574.

FENYÖ, E. M., MORFELDT-MÅNSON, L., CHIODI, F., LIND, B., VON GEGERFELT, A., ALBERT, J., OLAUSSON, E., and ÅSJÖ, B. (1988). Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. *J. Virol.* 62, 4414–4419.

FOLEY, G. E., LAZARUS, H., FARBER, S., UZMAN, B. G., BOONE, B. A., and MCCARTHY, R. E. (1965). Continuous culture of human lymphoblasts from peripheral blood of a child with acute leukemia. *Cancer* 18, 522–529.

FREDRIKSSON, R., STÅLHANSKE, P., VON GEGERFELT, A., LIND, B., ÅMAN, P., RASSART, E., and FENYÖ, E. M. (1991). Biological characterization of infectious molecular clones derived from a human immunodeficiency virus type-1 isolate with rapid/high replicative capacity. *Virology* 181, 55–61.

FROST, E., and WILLIAMS, J. (1978). Mapping temperature-sensitive and host-range mutations of adenovirus type 5 by marker rescue. *Virology* 91, 39–50.

GAZDAR, A. F., CARNEY, D. N., BUNN, P. A., RUSSELL, E. K., JAFFE, E. S., SCHECHTER, G. P., AND GUCCION, J. G. (1980). Mitogen requirements for the in vitro propagation of cutaneous lymphomas. *Blood* 55, 409–417.

GRAHAM, F. L., and VAN DER EB, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52, 456–467.

GROENINK, M., FOUCHER, R. A. M., DE GOEDE, R. E. Y., DE WOLF, F., GRUTERS, R. A., CUYPERS, H. T. M., HUISMAN, H. G., and TERSMETTE, M. (1991). Phenotypic heterogeneity in a panel of infectious molecular human immunodeficiency virus type 1 clones derived from a single individual. *J. Virol.* 65, 1968–1975.

HAHN, B. H., SHAW, G. M., ARYA, S. U., POPOVIC, M., GALLO, R. C., and WONG-STAAL, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature* 312, 166–169.

HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* 166, 557–580.

ISH-HOROWICZ, D., and BURKE, J. F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Res.* 9, 2989–2998.

LASKEY, L. A., NAKAMURA, G., SMITH, D. H., FENNIE, C., SHIMASAKI, C., PATZER, E., BERMAM, P., GREGORY, T., and CAPON, D. J. (1987). Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 50, 975–985.

LEONARD, J., PARROTT, C., BUCKLER-WHITE, A. J., TURNER, W., ROSS, E. K., MARTIN, M. A., and RABSON, A. B. (1989). The NF-κB binding sites in the human immunodeficiency virus long terminal repeat are not required for virus infectivity. *J. Virol.* 63, 4919–4924.

LEVY, J. A., CHANG-MAYER, C., DINA, D., and LUCIW, P. (1986). AIDS retrovirus (ARV-2) clone replicates in transfected human and animal fibroblasts. *Science* 232, 998–1001.

LUCIW, P. A., POTTER, S. J., STEIMER, K., DINA, D., and LEVY, J. A. (1984). Molecular cloning of AIDS-associated retrovirusvirus. *Nature* 312, 760–763.

MANN, D. L., O'BRIEN, S. J., GILBERT, D. A., REID, Y., POPOVIC, M., READ-CONNOLE, E., GALLO, R. C., and GAZDER, A. F. (1989). Origin of the HIV-susceptible human $CD4^+$ cell line H9. *AIDS Res. Human Retroviruses* 5, 253–255.

MOUNTS, P., WU, T.-C., and PEDEN, K. (1989). Method for cloning single-stranded oligonucleotides in a plasmid vector. *BioTechniques* 7, 356–359.

MUESING, M. A., SMITH, D. H., CABRADILLA, C. D., BENTON, C. V., LASKEY, L. A., and CAPON, D. J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 313, 450–458.

MYERS, G., BERZOFSKY, J. A., KORBER, B., SMITH, R. F., and PAVLAKIS, G. N. (1991). "Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences." Los Alamos National Laboratory, Los Alamos, NM.

RATNER, L., HASELTINE, W. A., PATARCA, R., LIVAK, K. J., STARCICH, B., JOSEPHS, S. J., DORAN, E. K., RAFALSKI, J. A., WHITEHORN, E. A., BAUMEISTER, K., IVANOFF, L., PETTEWAY, S. R., JR., PEARSON, M. L., LAUTENBERGER, J. A., PAPAS, T. S., GHRAYEB, J., CHANG, N. J., GALLO, R. C., and WONG-STAAL, F. (1985). Complete nucleotide sequence of AIDS virus, HTLV-III. *Nature* 313, 277–284.

SANCHEZ-PESCADOR, R., POWER, M. D., BARR, P. J., STEIMER, K. S., STEMPIEN, M. M., BROWN-SHIMER, S. L., GEE, W. W., RENARD, A., RANDOLPH, A., LEVY, J. A., DINA, D., and LUCIW, P. A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227, 484–492.

SANGER, F., NICKLEN, S., and COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

SCHUITEMAKER, H., KOOTSTRA, N. A., DE GOEDE, R. E. Y., DE WOLF, F., MIEDEMA, F., and TERSMETTE, M. (1991). Monocytotropic human

immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. *J. Virol.* 65, 356–363.

SMITH, S. D., SHATSKY, M., COHEN, P. S., WARNKE, R., LINK, M. P., and GLADER, B. E. (1984). Monoclonal antibody and enzymatic profiles of human malignant T-lymphoid cells and derived cell lines. *Cancer Res.* 44, 5657–5660.

SONIGO, P., ALIZON, M., STASKUS, K., KLATZMAN, D., COLE, S., DANOS, O., RETZEL, E., TIOLLAIS, P., HAASE, A., and WAIN-HOBSON, S. (1985). Nucleotide sequence of the Visna lentivirus: Relationship to the AIDS virus. *Cell* 42, 369–382.

SUNDSTRÖM, C., and NILSSON, K. (1976). Establishment and characterization of a human histiocytic lymphoma cell line (U-937). *Int. J. Cancer* 17, 565–577.

TERSMETTE, M., DE GOEDE, R. E. Y., AL, B. J. M., WINKEL, I. N., GRUTERS, R. A., CUYPERS, H. T., HUISMAN, H. G., and MIEDEMA, F. (1988). Differential syncytium-inducing capacity of human immunodeficiency virus isolates: Frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. *J. Virol.* 62, 2026–2032.

VARTANIAN, J.-P., MEYERHANS, A., ÅSJÖ, B., and WAIN-HOBSON, S. (1991). Selection, recombination, and G-A hypermutation of human immunodeficiency virus type I genomes. *J. Virol.* 65, 1779–1788.

WAIN-HOBSON, S., SONIGO, P., DANOS, O., COLE, S., and ALIZON, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* 40, 9–17.

WAIN-HOBSON, S., VARTANIAN, J.-P., HENRY, M., CHENCINER, N., CHEYNIER, R., DELALASSUS, S., PEDROZA-MARTINS, L., SALA, M., NUGEYRE, M.-T., GUÉTARD, D., KLATZMANN, D., GLUCKMAN, J.-C., ROZENBAUM, W., BARRÉ-SINOUSSI, F., and MONTAGNIER, L. (1991). LAV revisited: Origins of the early HIV-1 isolates from Institut Pasteur. *Science* 252, 961–965.

WEISS, A., WISKOCIL, R. L., and STOBO, J. D. (1984). The role of T3 surface molecules in the activation of human T cells: A two-stimulus requirement for IL-2 production reflects events occurring at a pre-translational level. *J. Immunol.* 133, 123–131.

WIGLER, M., PELLICER, A., SILVERSTEIN, S., AXEL, R., URLAUB, G., and CHASIN, L. (1979). DNA-mediated transfer of the adenine phosphoribosyltransferase locus into mammalian cells. *Proc. Natl. Acad. Sci. USA* 76, 1373–1376.

WILLEY, R. L., SMITH, D. H., LASKEY, L. A., THEODORE, T. S., EARL, P. L., MOSS, B., CAPON, D. J., and MARTIN, M. A. (1988). In vitro mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. *J. Virol.* 62, 139–147.

WILLEY, R. L., ROSS, E. K., BUCKER-WHITE, A. J., THEODORE, T. S., and MARTIN, M. A. (1989). Functional interaction of constant and variable domains of human immunodeficiency virus type 1 gp120. *J. Virol.* 63, 3595–3600.

Journal of Acquired Immune Deficiency Syndromes
2:344–352 © 1989 Raven Press, Ltd., New York

HIV-1 Isolates Are Rapidly Evolving Quasispecies: Evidence for Viral Mixtures and Preferred Nucleotide Substitutions

Maureen Goodenow, Thierry Huet, *William Saurin, †Shirley Kwok, †John Sninsky, and Simon Wain-Hobson

*Laboratoire de Biologie et Immunologie Moléculaires des Rétrovirus and *Unité de Programmation Moléculaires et Toxicologie Génétique, Institut Pasteur, Paris, France, and †Department of Diagnostic Research, Cetus Corporation, Emeryville, California*

Summary: RNA viruses are renowned for their genetic variability. The human immunodeficiency viruses (HIV) are no exception. A rapid method has been established for the genetic identification and differentiation of viral strains based on the sequencing of many M13 clones of gene-amplified products. Some isolates are internally relatively homogeneous while others are heterogeneous. There was no correlation between virus complexity and disease stage. One isolate was in fact a mixture of two distinct strains. A strong preference for G → A base substitutions was observed. These data indicate that HIV isolates cannot be described in simple molecular terms and should rather be considered as quasispecies. Key Words: HIV—Genetic variability—Viral mixtures—Preferred base substitutions—Quasispecies.

The ability of viral RNA genomes to undergo genetic change is in part due to the extensive misincorporation of nucleotides by the viral polymerase. While some RNA viruses seem to be relatively stable (e.g., measles and poliovirus), others are highly labile—influenza A being a case in point. Nucleotide misincorporation rates of greater than 10^{-4}/base/cycle have been reported (reviewed in ref. 1). The AIDS viruses HIV-1 and HIV-2 are genetically highly variable with the surface envelope protein (gp120) sequence being the most variable of all (2–5). The gp120 protein sequence is composed of interspaced hypervariable (HY) and constant (C) regions. The differences in the hypervariable regions are not restricted to amino acid substitutions but include duplications and deletions of small segments.

The task of characterizing HIV isolates and defining interrelationships is complicated by the lack of good antibody reagents. In fact, most neutralizing antibodies are type specific (6,7). Consequently, HIV-1 strain variability has been analyzed by genetic methods, notably by restriction mapping, which offer low resolution (8,9). Molecular cloning and sequencing of entire genomes gives better resolution but is time consuming (2). Rapid methods to type HIVs are needed. Such techniques could be exploited to answer a number of basic questions, such as is there any correlation between the viral strain and the clinical course of infection? Or what is the rate of change of the virus in vivo and in vitro? Or does superinfection occur? Given that the majority of seropositive individuals have had multiple

Address correspondence and reprint requests to Dr. S. Wain-Hobson at Laboratoire de Biologie et Immunologie Moléculaires des Rétrovirus, Institut Pasteur, 28 Rue du Dr. Roux, 75724 Paris cédex 15, France.

The present address of Dr. M. Goodenow is Department of Pathology, JH Miller Health Center, University of Florida, Gainesville, FL 32610, U.S.A.

partners or exposure to HIV-infected material, superinfection is a distinct possibility.

With these questions in mind, we have established a rapid genetic method for characterizing HIV-1 isolates using the polymerase chain reaction (PCR, ref. 10) to amplify HIV-1 DNA (11,12). *gag* and *env* segments were amplified, cloned into M13, and multiple recombinants were sequenced. Apart from being able to distinguish between different HIV isolates, it became clear that HIV-1 isolates were internally highly polymorphic and could be made up of two different viruses. Thus, HIV-1 cannot be described in simple molecular terms and should be considered rather as quasispecies.

METHODS

Viral DNA and PCR Amplification

High molecular weight DNA was extracted from peripheral blood mononuclear cells infected by cocultivation. HIV subgenomic regions were amplified by PCR. The *gag* region primers were SK38 and SK39 (ref. 12) while for the *env* region SK122 (5′ CAAGCCTAAAGCCATGTGTA 3′) and SK123 (5′ TAATGTATGGGAATTGGCTCAA 3′) were used. Thirty cycles of amplification were performed in a Perkin Elmer–Cetus Thermal Cycler as recommended. The hybridization, polymerization, and denaturing temperatures were 55, 70, and 95°C, respectively. The melting temperatures of SK38, SK39, and SK19 oligonucleotides to template DNA in 2.5 m*M* Mg^{2+} were measured. They were 65, 68, and 67°C, respectively, and in good agreement with the calculated values of 67, 72, and 72°C. The calculated values for SK122, SK123, and SK129 are 66, 67, and 68°C, respectively. Therefore, working at a hybridization temperature of 55°C, approximately 10°C below the T_m of the primers, indicated that they should have been able to prime from some heterologous sequences. Approximately 10% of the amplified material was blunt end ligated directly into HincII cleaved M13mp10w and transformed into DG98. M13 plaques were screened with ^{32}P-labeled oligonucleotides that mapped within the amplified regions. The *gag* and *env* probes were SK19 (12) and SK129 (5′ TGTAAAATTAACCCCACTCTGTGTTA 3′), respectively. After purification, phages were sequenced by the dideoxy method using [α-^{35}S]dATP labeling and buffer gradient gels (13).

Tree Reconstruction and Counting of Mutational Events

For each isolate, the tree topology was determined by which the different sequences could be either at nodes (ancestries) or at leaves. Trees were computed that implied the minimum number of changes between sequences using a two-step method. First, Prim's algorithm (14) was used. This algorithm finds minimal trees from dissimilarity measures between elements of a set. In our case, the dissimilarity measure was the number of differences in a pairwise comparison between sequences within a given isolate. Because of the possibility of multiple mutations at a given site and due to the slight length variation within some sets of data, such trees are not necessarily minimal. Trees with minimal length were selected taking into account all possible point mutations. The different kinds of mutations were then counted, assuming that the ancestor sequence in the isolate was the most abundant. When more than one minimal tree was found, the number and kind of mutation was calculated as the average of the numbers in each possible tree.

RESULTS

P25*gag* Region

Of the seven viral isolates analyzed, six have not been previously studied by genetic methods. HIV-1 isolates BAN and TRA were derived from two French AIDS patients; B88, typical of a slow–low virus, and B40, a rapid–high virus (15), were isolated from Swedish patients with persistent generalized lymphadenopathy and AIDS, respectively; the MAB isolate was from a Gabonese patient with frank AIDS while isolate 397 was from a healthy Gabonese presenting an atypical Western blot (T. Huet et al., unpublished data). The seventh isolate, HIV-1 ELI (Zaïre), was included as a positive control since a complete provirus had been previously cloned and sequenced from the same DNA preparation (2). All isolates had been propagated only on peripheral blood mononuclear cells for a limited number (<7) of passages.

In order to assess internal variability, two regions of the HIV-1 genome were analyzed. The first was a small, highly conserved region of 59 bases (12,16) in p25*gag*. The combined nucleic acid and protein sequence data are given in Fig. 1. Every isolate was heterogeneous, harboring from 3–7 genomes. Usu-

VIRUS. CLONE	NUCLEIC ACID SEQUENCE	FREQUENCY NA	FREQUENCY AA	AMINO ACID SEQUENCE
MAB.102	CTATAAAAGATGGATAATCCTGGGGTTAAATAAAATAGTAAGAATGTATAGCCCTGTCA	61%	74%	YKRWIILGLNKIVRMYSPV
MAB.105	T A	13%		: :
MAB.103	G A	13%	13%	V :
MAB.101	T A AC	13%	13%	: T
BAN.101	CTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACTA	76%	92%	YKRWIILGLNKIVRMYSPT
BAN.117	C	8%		:
BAN.112	A	4%		:
BAN.114	____ G C	4%		: :
BAN.106	A	4%	4%	M
BAN.118	A	4%	4%	E
B88.101	CTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCA	85%	95%	YKRWIILGLNKIVRMYSPT
B88.106	A	10%		:
B88.105	A	5%	5%	K
ELI.102	CTATAAAAGATGGATAATTCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTGTCA	88%	100%	YKRWIILGLNKIVRMYSPV
ELI.115	A	8%		:
ELI.101	C	4%		:
TRA.107	CTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCA	66%	88%	YKRWIILGLNKIVRMYSPT
TRA.101	G T	22%		: :
TRA.108	C	4%	4%	T
TRA.109	G	4%	4%	M
TRA.113	G	4%	4%	C
397.101	CTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCA	75%	80%	YKRWIILGLNKIVRMYSPT
397.114	T	5%		:
397.111	A	5%	5%	K
397.107	A	10%	15%	*
397.110	AA	5%		*
B40.101	CTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTATCA	44%	56%	YKRWIILGLNKIVRMYSPI
B40.105	T	4%		
B40.122	T C	4%		:
B40.112	T G	4%		:
B40.103	T C G C	31%	40%	: : T
B40.109	C G C	9%		: : T
B40.121	T A A C G C	4%	4%	* : : : T

FIG. 1. Nucleotide and amino acid variability within the p25*gag* region. Only those clones representing a unique sequence are given. The frequencies (NA, nucleic acid; AA, amino acid) refer to the proportion of M13 subclones sequenced having an identical sequence. In all cases, the sequence shown is that of the most abundant form while only the substitutions are given for the other clones. Within the amino acid data, an asterisk in the sequence (e.g., 397.107, 397.110, and B40.121) denotes a stop codon. A colon represents a silent nucleotide change within a codon. The reading frame is picked up at position two. Thus, the first and last bases do not figure among the peptide data. The numbers of clones sequenced were MAB, 8; BAN, 25; B88, 20; ELI, 23; TRA, 23; 397, 20; and B40, 23. Protein sequences are given using the one letter code: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

ally, a particular genome predominated (60–85%), with others differing by no more than 1–2 bases, present at frequencies of 4–22%. The Swedish isolate B40 proved to be different. Two sequences, B40.101 and B40.103, were present at comparable frequencies of 44 and 31%, respectively, with minor forms of each present at 4–9%. The two sequences differed by four nucleotides (4/59 = 7%), which is comparable to the difference between the corresponding sequences of HIV-1 strain BRU (France)

and either of strains CDC451 (USA) or MAL (Zaïre). Thus, it would seem that the B40 isolate is composed of two viruses.

Only point mutations (with respect to the dominant genome sequence) were observed. Of the amino acid substitutions, a number were nonconservative (e.g., Leu → Glu or Tyr → Cys). Unexpectedly, the single tryptophan codon (TGG) was found mutated to an opal (TGA, 397.107), an ocher (TAA, 397.110), and an amber (TAG, B40.121) stop codon (Fig. 1). In the case of isolate 397, 15% (3/20) of sequences carried a stop codon. Since this segment represents a mere 0.6% of the HIV genome, these data would suggest that virtually all of the HIV-1 397 genomes are multiply defective.

In view of such complexity, the word "genome" will be used to represent a unique viral nucleic acid sequence, "strain" will represent a cluster or set of highly related genomes, and "isolate" the resulting collection of genomes derived from a seropositive individual.

First *env* Hypervariable Region

The first hypervariable region (HY1) was chosen for a number of reasons: (a) each isolate hitherto studied had given a distinct sequence; (b) it carried a number of direct repeats; (c) it was approximately 300 bases long, indicating that it could be readily sequenced using a single buffer gradient gel; and (d) it was surrounded by highly conserved regions in which to anchor the amplification primers.

The data from this region of six of the viruses are shown in Fig. 2. Only the protein data have been given since the nucleic acid sequence data (28 kb) were too voluminous. The data are available upon request. What became apparent was that each virus represented a distinct set of sequences of varying degrees of complexity. Thus, the HIV-1 BAN isolate (France) was homogeneous at the protein sequence (BAN.12 carried two silent mutations) while the HIV-1 B88 (Sweden) isolate had a single amino acid substitution at position 16.

For HIV-1 ELI (Zaïre), two genomes ELI.01 and ELI.04, differing by a single Arg → Gly substitution, accounted for 86% of the genomes sequenced. Clone ELI.03 proved to be very different. Firstly, it was yet another example of a defective genome but this time a single A nucleotide had been inserted within a run of As at codon 32. Secondly, it differed by 18/274 (7%) bases with respect to ELI.01. It could represent either a trace of a second virus or simply a defective genome accumulating mutations faster.

HIV-1 TRA (France) proved more complex. The dominant genome, TRA.01, was present at only 40% while most of the other sequences were clustered about this sequence and differed by only 1–4 bases. TRA.04 was the first case of a defective genome due to a single base deletion. Two additional forms (TRA.18 and TRA.11) differed by 10 and 11 bases, respectively, with respect to the major genome sequence. It should be noted that two-thirds of the point mutations resulted in amino acid substitutions.

More complex still was the HIV-1 397 isolate (Gabon). Here a single form predominated, with 10 minor forms representing a tight set of genomes. Three points could be made: (a) all but one of the 12 point mutations resulted in amino acid substitutions; (b) internal length heterogeneity was due to deletion of 33 bases on two separate occasions (e.g., 397.09 and .06); this could be due to slippage between the direct repeats (underlined) in the protein sequence VNTTSSSLRNATNTTSSS during reverse transcription; and (c) in the case of 397.06, the deletion occurred between valine codon 14 (G/TT) and threonine codon 25 (A/CT), giving rise to a deletion of 11 residues and a novel alanine (G/CT) codon.

The HIV-1 B40 (Sweden) *env* data could be resolved into two clusters of genomes, confirming that the B40 isolate was a mixture of two distinct viruses. One was defined by the B40.02 sequence while the second virus was represented by a disperse cluster of genomes, B40.01, B40.08, and B40.14. It can be seen from Fig. 2 that each virus has a unique sequence length for this region varying from 83 residues (BAN) to 95 (397). The length difference between B40.02 and B40.01 in conjunction with the ten base differences between all three and B40.02 supported the earlier observation from the *gag* data that in fact the B40 was a mixture of two distinct viruses.

The PCR method depends on in vitro polymerization at elevated temperatures. It was therefore possible that a proportion of the base substitutions and insertions/deletions were due to the Taq polymerase. Saiki et al. (ref. 17) have estimated the error rate as 1/400 bases sequenced per 30 cycles. Tindall and Kunkel (18), using a different system, have estimated the single base substitution and deletion error rates to be 1.1×10^{-4} and 2.4×10^{-5}, respectively. To control for possible errors due to

VIRUS. CLONE	FREQUENCY NA	AA	AMINO ACID SEQUENCE
			10 20 30 40 50 60 70 80 83
BAN.01	89%	100%	KLTPLCVTLN CTKLKNVTYA NNSSREKEEM KNCSFTTRIG NKVQKEYALF YKLDVVPIDK DNTSFILIHC NTSTITQACP KVS
BAN.12	11		: :
			10 20 30 40 50 60 70 80 87
B88.01	95%	95%	KLTPLCVTLN CTDGLRNATN NNDSSWGGEM KNCSFNITTS IRDKVQKEYA LFYKLDVVPI DSDNNSTKYR LINCNTSVIT QACPKIS
B88.02	5	5	K
			10 20 30 40 50 60 70 80 90 91
ELI.01	50%	57%	KLTPLCVTLN CSDELRNNGT MGNNVTTEEK GMKNCSFNVT TVLKDKKQQV YALFYRLDIV PIDNDSSTNS TNYRLINCNT SAITQACPKV S
ELI.11	7		:
ELI.10	7	7	K
ELI.04	29	29	G :
ELI.03	7	7	NK:K IE K K° N: K N N K
			10 20 30 40 50 60 70 80 89
TRA.01	40%	40%	QLTPLCVTLN CTDYLGNATN TTSSSGGGME RGEIKNCSFN ITTSIRDKMQ REYALFYKLD VVPTDNDNTS YRLISCNTSV ITQACPKVS
TRA.06	5	10	R
TRA.17	5		R : : :
TRA.19	5	5	N
TRA.15	5	5	S
TRA.11	10	10	: KK IA : : :
TRA.18	5	5	S RRR K K KK K T //////
TRA.08	10	15	I
TRA.14	5		I :
TRA.04	5	5	I •
			10 20 30 40 50 60 70 80 90 95
397.02	56%	56%	KLTPLCVTLD CTDVNTTSSS LRNATNTTSS SWETMEKGEI KNCSFNTTTS IRDKMQEQYA LFYKLDVLPI DKNDTKFRLI HCDTSTITQA CPKIS
397.04	4	4	
397.12	4	4	T
397.10	4	4	I
397.03	4	4	N
397.05	4	4	N : S //
397.18	4	4	S S
397.27	4	4	---------- -
397.11	4	4	---------- - D
397.09	4	4	---------- - P
397.06	4	4	I -------- ----A N
			10 20 30 40 50 60 70 80 88
B40.02	42%	67%	KLTPLCVTLN CTDLRNDTNT NSSSWGKMER GEIKNCSFNV TTSIRNKMQK EYALLYKLDV VPIDNDTTSY RLISCNTSVI TQACPKVT
B40.06	17		:
B40.05	8		:
B40.03	8	8	L
B40.01	8	8	: K ---A G M K M: D V :G
B40.08	8	8	: K ---A G M K M: D V K : A R
B40.14	8	8	: K ---A G M K M: D V

FIG. 2. Amino acid sequence variation within the first hypervariable region (HY1) of *env* gp120. Six isolates were studied in detail, resulting in 28 kb of data. Only protein sequences are given. Nucleic acid sequences are available on a PC-compatible floppy disc. The frequencies are as in Fig. 1. Only amino substitutions with respect to the most abundant sequence are given. Dashes represent gaps introduced to maximize homology. The symbol ° denotes a frameshift mutation, either a single base deletion (TRA.04 and 397.02) or a single base insertion (ELI.03). A colon represents a silent nucleotide change within a codon. Slashes represent the premature end of an individual clone due to small deletions of a few bases that occurred only at the 3' end. The number of clones sequenced was BAN, 9; B88, 19; ELI, 12; TRA, 18; 397, 23; and B40, 12.

Taq, the following experiments were performed: a plasmid carrying the relevant HIV-1 SF2 *gag* sequence was diluted to approximately 10^4 copies of target in 1 μg of high molecular weight uninfected human lymphocyte DNA (equivalent of 150,000 cells). After 30 cycles of PCR using the same *gag* primers, the products were cloned and 19 M13 recombinants were sequenced. Not a single substitution or deletion was noted (data not shown). For the *env* gene, which carried large deletions among some clones of isolate 397, DNA from recombinant clone 397.02 was diluted and amplified. A total of 15 M13 subclones were sequenced. Again, neither substitutions nor deletions/insertions were detected (data not shown). It is therefore possible to calculate the minimal error frequency in the to be $<1/5,411$ bases

sequenced after 30 cycles of PCR, a Taq polymerase error rate of $<1.5 \times 10^{-5}$. In addition the decrease in nucleotide triphosphate and magnesium concentrations led to a 10-fold better error rate (7). Thus, for our study of sets of 15–25 *gag* or *env* sequences, essentially none of the observed differences represented artifacts.

Biased Substitution Frequencies

A cursory glance at the *gag* and *env* data indicated an inordinate number of G → A transitions relative to the viral (+) strand. This was examined by a phylogenetic tree analysis based on the assumption that the minor forms were derived from the major forms by the least number of mutations (maximum parsimony). Once the phylogenetic relationships had been established, it was possible to describe the minimum number of mutations linking all of the sequences to the major form. The minimum number of mutations for all the *env* and *gag* data sets is presented in Table 1. This table gives the substitution frequencies for a given base (*N*) per thousand *N* bases. This normalization was necessary since the composition of the HIV genome is particularly A rich. Thus, for the BRU strain of HIV-1, the base composition is 22.2% T, 17.8% C, 35.8% A, and 24.2% G (ref. 19). In fact, the most common base change observed within both sets of data was indeed the G → A transition. This transition was 4- to 15-fold greater than any other transition. The frequency of the reciprocal transition, A → G, was comparable to those for the C → T or T → C transitions. Transitions (R ⇐ R or Y ← Y) were approximately 15-fold more frequent than transversions (R ↔ Y). When normalized to the base composition per codon base (first, second, and third positions), it was clear that for the *env* gene this preference was independent of the position within the codon, arguing against any strong selective pressure being operative on this region. There were fewer substitutions in the *gag* data set. For this reason, only the overall substitution frequencies have been given. Nonetheless, the same preference for G → A transitions was observed although it was somewhat less intense. The observation that for virtually every data set the G → A substitution was dominant argues in favor of the assumption that the minor forms were derived from the major form.

TABLE 1. *Biased nucleotide substitution frequencies among the clusters of HIV-1 sequences*

	Mutational events	Position in codons I	II	III	Mean
env	A → N	2.9	7.6	3.5	4.5
	G → N	33.0	30.9	25.9	30.4
	C → N	6.7	4.6	1.9	4.0
	U → N	1.7	5.1	7.4	5.2
	All	9.5	10.7	7.8	9.4
	Mutational events	Position in codons I	II	III	Mean
gag	A → N	3.2	7.6	23.8	11.2
	G → N	11.4	38.5	82.1	45.9
	C → N	19.4	0.0	32.5	21.4
	U → N	14.8	15.7	31.6	19.9
	All	8.0	18.2	36.2	20.9

The number of mutational events ($X \rightarrow N$) are given per thousand sites having a changed nucleotide (X) in the *gag* and *env* data sets. The numbers of mutations are given for each position in the codons (I, II, and III) as well as for the entire sequence (mean). *env:* there is no statistical difference between the normalized substitution frequencies for each codon position. The extraordinary high G → N frequencies with respect to the other changes ($\chi^2 = 90.16$, 3 degrees of freedom) are dominated by the G → A transition. The numbers of changes affecting the three other bases are not significantly different. The ratio of transitions to transversions is 15:2. *gag:* this region shows the higher variability typical of the third base with respect to positions I and II ($\chi^2 = 10.36$, 2 degrees of freedom). The numbers of changes affecting G is greater than any other base ($\chi^2 = 11.34$, 3 degrees of freedom). The ratio of transitions to transversions is 15:1. For both regions, no nearest-neighbor effect around the sites of G → A transitions could be detected. The assumption that the minor forms are derived from the major forms is probably good since in virtually every data set the G → A preference was always dominant.

The nature of the substitutions was not the same for the *env* and *gag* data sets. Within the *gag* data, there were 7 expressed and 12 silent codon changes while for the *env* data 52 expressed and 17 silent codon changes were noted. Thus, despite manifesting comparable substitution frequencies, there appear to be some constraints on the *gag* region that are more important than for the HY1 region. Finally, it should be noted that the three frameshift mutations identified among the *env* data sets involved the insertion (once) or the deletion (twice) of a single A residue within runs of A residues.

DISCUSSION

RNA viruses have been described as quasispecies (1,20,21), i.e., there is no such thing as a viral sequence per se but sets or clusters of closely related sequences. Such is the case for HIV. From any set of data that has been derived, it was simple to calculate that every HIV viral genome within an isolate was unique. This in turn meant that the rate of nucleotide misincorporation was greater than 1×10^{-4}/base/cycle of replication. This value is comparable to that established for Rous sarcoma virus (1.4

$\times 10^{-4}$; ref. 22) and the vesicular stomatitis virus L gene ($1–4 \times 10^{-4}$; ref. 23) and greater than that of the influenza A virus NS protein (1.5×10^{-5}; ref. 24) and polio VP1 protein (2.1×10^{-6}; ref. 24). The potential of the HIV-1 virus to change is thus enormous but not necessarily greater than that of some other RNA viruses. The important difference between HIV and other RNA viruses is its ability to persist despite a vigorous immune response. Thus, it would seem that, like visna virus (25), variants are accumulated in vivo.

The genetic variability within these isolates has probably been underestimated for methodological reasons. (a) The choice of primers and probes restricted the population of sequences amplified. Although working at a hybridization temperature of 55°C, approximately 10°C below the T_m of the primers (see the legend to Fig. 1), some highly divergent genomes would not have been detected. In addition, there remains the possibility that the penultimate or most 3′ base of the primer is not complementary to the majority of target sequences. In this case, any amplified DNA might represent only a minor form! (b) Minor forms (<4–5%) went undetected since our analysis was limited to 15–25 M13 subclones per isolate with the exception of two, for which 8 and 9 subclones per isolate were sequenced. Thus, the isolates will appear to be more homogeneous than they probably were.

With these reservations in mind, it has been possible to show that viral isolates can harbor diverse sequences (e.g., isolates ELI and TRA) and can even be mixtures of two different strains (isolate B40). We cannot rigorously conclude that this is evidence of superinfection or coinfection in vivo due to the unfortunate ease by which cultures can become contaminated (26).

The observation of defective genomes was intriguing. While not new in retrovirology, a frequency of 15% at a single site in the HIV-1 397 strain was surprising. Were this to be typical of every region of the 397 genome, it is simple to calculate that every genome must be multiply defective. Clearly, this is not the case for all isolates since a number of infectious molecular clones exist (27). The presence of some in phase terminator codons must not be overinterpreted since Moloney murine leukemia virus is capable of upregulating host cell suppressor tRNA genes (28).

Given the intense sequence variability within the HY1 region, such that within just a short segment of 80–100 amino acids two different viruses can be distinguished, we now have a genetic method for "quality control," strain verification, and subtyping of isolates.

Quality Control

It is evident that isolates BAN and B88 represent relatively good reagents for viral envelope studies like infectivity or neutralization. Isolates like TRA or B40 would have to be plaque purified before embarking on any detailed analysis. Furthermore, by repeated analyses, it would be possible to find out if a virus changed in vitro.

Strain Verification

By regular amplification and direct sequencing of the products, it can be seen whether or not an isolate has become contaminated by another, perhaps faster growing or better adapted virus. This approach would have very quickly revealed the true identity of the so-called STLV-3-agm and HTLV-4 viruses, which were contaminants of the SIV-mac251 strain (26). The advantage of cloning the amplified products and sequencing 10–20 clones is that the sequence dispersion within a given isolate may be established, so eliminating the reliance on a single molecular clone.

Subtyping

This application is just a different facet of the above point. Having a powerful molecular method available, it will be possible to seek correlations between strains and disease progression (or speed thereof), intensity of the immune responses, and geography. Another application could be to assess the role of superinfection in the development of disease. Already, it is clear that there is no simple correlation between the heterogeneity of a given isolate and the clinical stage of the patient at the time of virus isolation (e.g., BAN and B40 or B88 and 397).

That the large number of base substitutions observed were due to the Taq polymerase has been discounted experimentally. Furthermore, if the G $\rightarrow$ A bias, for example, were due to Taq polymerase misincorporation, it would have been symmetric. Consequently, reciprocal base substitution frequencies would be equal (i.e., G $\rightarrow$ A = A $\rightarrow$ G), which is clearly not the case. Incidentally, when Taq polymerase does misincorporate nucleotides, a totally

different order is seen (T → C > C → T ~ A → G ~ G → A; ref. 18).

The bias towards G → A transitions on the (+) strand suggests that they arose in one of the two highly asymmetric steps in HIV replication. These are the reverse transcription of RNA into the (−) strand DNA and the transcription of the proviral (−) DNA strand into RNA by RNA polymerase II. In order to produce a G → A transition, either a rG:dT mismatch must be introduced into the RNA–DNA heteroduplex during reverse transcription or else a rA:dC mismatch occurs in the RNA–DNA heteroduplex during transcription of the provirus. G:T and G:U mismatches, respectively, in DNA and RNA, are known to be less destabilizing than A:C pairs (29–31). In addition, reverse transcriptase may have a reduced specificity since it is both a RNA- and a DNA-dependent DNA polymerase whereas RNA polymerase has a unique specificity for DNA templates. Finally, it is interesting to note that the most frequent reverse transcriptase errors picked up by sequencing multiple cDNA clones are G → A substitutions and insertions or deletions in runs of adenosines (D. Capon, personal communication).

This G → A bias helps us to understand the high A content (34%) of HIV. There are two corrolaries. Firstly, since the frequency of G → A substitution is greater than A → G, there will be an inexorable trend towards an even more A-rich genome (a drive to polyA), meaning that HIV is not in equilibrium. Secondly, loci or codons particularly G rich will be particularly unstable. Therefore, it was perhaps not surprising to have observed stop codons (TGA, TAA, and TAG) derived from a tryptophan codon (TGG). More importantly, it helps to explain why the hypervariable regions are variable. These regions are in contact with solvent and are generally polar in composition (Asn, Lys, Asp, Glu, Arg, and Ser). Since many of their corresponding codons are G rich, these sequences are thus most vulnerable to change. Amino acid substitutions will be polar, thus not changing the nature of these hypervariable regions.

In conclusion, we are faced by a virus of enormous complexity, certainly more heterogeneous than influenza A or poliovirus (24). The data described here suggest that there may be as many viral strains of HIV-1 as there are carriers. Given this background, the striking similarity between the LAV-1 and HTLV-3B strains of HIV-1 remains a signal exception. The possibility of viral mixtures within isolates as well as the high frequency of apparently defective genomes render the task of the molecular biologist difficult. The impact of such diversity upon the immune system has to be defined. Only by concentrating on highly conserved regions such as active site residues might it be possible to develop general strategies for controlling HIV infection.

Acknowledgment: We would like to thank Dr. Brigitta Asjö (viruses B88 and B40), Dr. Françoise Brun-Vézinet (viruses TRA and BAN), Dr. Eric Delaporte (viruses 397 and MAB), and Dr. Luc Montagnier (virus ELI) for gifts of their HIV-1 viral strains; Drs. Steve Isaacs and George Ciminohr of HR1 Inc. for the determination of the Tms of the gag primers; Dr. Tom Kunkel for making available preprints of his work; and Drs. Philippe Marlière, Andreas Meyerhans, and Fernando Plata for critical suggestions. Dr. Goodenow was supported by a travelling fellowship from l'Institut National de la Santé et de la Recherche Médicale; Thierry Huet was supported by a grant from Institut Merieux. This work was supported with grants from Institut Pasteur and La Fondation pour la Recherche Médicale, and l'Association pour la Recherche sur le Cancer.

REFERENCES

1. Steinhauer DA, Holland JJ. Rapid evolution of RNA viruses. *Annu Rev Microbiol* 1986;41:409–33.
2. Alizon M, Wain-Hobson S, Montagnier L, Sonigo P. Genetic variability of the AIDS virus: nucleotide sequence of two isolates from African patients. *Cell* 1986;46:63–74.
3. Starcich B, Hahn B, Shaw G, McNeely P, et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 1986;45:637–48.
4. Willey RL, Smith DH, Lasky LA, et al. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc Natl Acad Sci USA* 1986;83:5038–42.
5. Guyader M, Emermann M, Sonigo P, Clavel F, Montagnier L, Alizon M. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature (Lond)* 1986;326:662–8.
6. Weiss RA, Clapham PR, Weber JN, Dalgleish A, Laskey LA, Berman PW. Variable and conserved neutralization antigens of HIV. *Nature (Lond)* 1986;324:572–5.
7. Nara P, Robey WG, Pyle SW, et al. Purified envelope glycoproteins from human immunodeficiency virus type 1 variants induce individual, type-specific neutralizing antibodies. *J Virol* 1988;62:2622–8.
8. Benn S, Rutledge R, Folks T, et al. Genomic heterogeneity of AIDS retroviral isolates from North America and Zaire. *Science* 1985;230:949–51.
9. Saag MS, Hahn BH, Gibbons J, et al. Extensive variation of human immunodeficiency virus type-1 in vivo. *Nature (Lond)* 1988;334:440–4.
10. Saiki R, Sharf S, Faloona F, et al. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 1985;230:1350–4.
11. Kwok S, Mack DH, Mullis KB, et al. Identification of human immunodeficiency virus sequences by using in vitro

enzymatic amplification and oligomer cleavage detection. *J Virol* 1987;61:1690–4.

12. Ou CY, Kwok S, Mitchell SW, et al. DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells. *Science* 1988;239:295–7.
13. Biggin MD, Gibson TJ, Hong GF. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc Natl Acad Sci USA* 1983;80:3963–5.
14. Prim RC. *Bell Syst Tech J* 1957;36:1389–401.
15. Asjō B, Morfeldt-Manson L, Albert J, et al. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. *Lancet* 1986;2:660–2.
16. Myers G, Rabson AB, Josephs SF, Smith TF, Wong-Staal F. Human retroviruses and AIDS, a compilation and analysis of nucleic acid and amino acid sequences. *Human retroviruses and AIDS*. Los Alamos: Los Alamos National Laboratory, 1988.
17. Saiki R, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;239:487–91.
18. Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the thermus aquaticus DNA polymerase. *Biochemistry* 1988;27:6008–13.
19. Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. *Cell* 1985;40:9–17.
20. Eigen M. Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 1973; 58:65–523.
21. Domingo E, Martinez-Salas E, Sobrino F, et al. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* 1985;40:1–40.
22. Leider JM, Palese P, Smith FI. Determination of the mutation rate of a retrovirus. *J Virol* 1988;62:3084–91.
23. Steinhauer DA, Holland JJ. Direct method for quantification of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* 1986;57:219–28.
24. Parvin JD, Moscona A, Pan WT, Leider JM, Palese P. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J Virol* 1986;59:377–83.
25. Lutley R, Petursson G, Palsson PA, Georgsson G, Klein J, Nathanson N. Antigenic drift in visna: virus variation during long term infection of Icelandic sheep. *J Gen Virol* 1983;64:1433–40.
26. Kestler HW, Li Y, Naidu YM, et al. Comparison of simian immunodeficiency virus isolates. *Nature* (*Lond*) 1988; 331:619–21.
27. Fisher AG, Feinberg MB, Josephs SF, et al. The transactivator gene of HTLV-III is essential for virus replication. *Nature* (*Lond*) 1986;320:367–71.
28. Yoshinaka Y, Kathoh I, Copeland TD, Oroszlan S. Murine leukemia virus protease is encoded by the gag–pol gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci USA* 1985;82:1618–22.
29. Tinoco I, Borer PN, Dengler B, et al. Improved estimation of secondary structure in ribonucleic acids. *Nature* (*New Biol*) 1973;246:40–1.
30. Ikuta S, Takagi K, Wallace RB, Itakura K. Dissociation kinetics of 19 base paired oligonucleotide–DNA duplexes containing different single mismatched base pairs. *Nucl Acids Res* 1987;15:797–811.
31. Modrich P. DNA mismatch correction. *Annu Rev Biochem* 1987;56:435–66.

Rapid generation of sequence variation during primary HIV-1 infection

Shen Pang, Yshay Shlesinger, Eric S. Daar*, Tarsem Moudgil*, David D. Ho† and Irvin S.Y. Chen‡

Objective: HIV-1 undergoes extensive genetic variation in infected individuals. The extent of genetic variation has been examined in patients with AIDS, but little is known regarding the appearance of HIV-1 genetic variation immediately following infection during the primary phase of HIV-1 infection prior to seroconversion.

Design: We examined HIV-1 genetic variation during this early phase of HIV-1 infection by polymerase chain reaction (PCR) and nucleotide sequence analysis of the V4 variable region and the CD4-binding domain.

Results: Our results demonstrate that extensive sequence variation is seen early after infection, although a predominant HIV-1 species is maintained.

Conclusions: The type of variants that occur are dynamic, changing over time, and the mutations seen are consistent with those expected from random occurrence, unlike the pattern of variation previously reported during later stages of disease.

AIDS 1992, 6:453–460

Keywords: HIV-1 DNA, genetic variation, mutation, seroconversion.

Introduction

It is recognized that the HIV-1 genome exhibits tremendous sequence variation among different HIV-1 strains, as well as within a given individual [1–6]. Generally, the nucleotide sequence variation of HIV-1 genomes within an individual is less than that among HIV-1 strains from different individuals, suggesting evolution within an individual from a limited number of infecting HIV-1 virions. The collection of distinct but related variants within an individual has been termed 'quasi-species' [1,7]. The pattern of nucleotide changes suggests that selection pressure drives the appearance of certain variants, since the frequency of silent mutations is much lower than would be predicted on a random basis [8]. A number of questions remain unanswered regarding the rate at which variants emerge *in vivo* and the nature of the selection processes that may be important for sequence variation and, ultimately, viral pathogenesis.

We have addressed some of these issues by examining HIV-1 sequence variation shortly after *in vivo* infection.These primary or acute cases of HIV-1 infection can be identified by influenza-like symptoms occuring on average 6–12 weeks after exposure to HIV-1, concurrent with development of HIV-1-specific antibodies. Acute HIV-1 infection in an HIV-1-antibody-negative individual can be confirmed by detection of p24 antigen, culturing and/or by polymerase chain reaction (PCR). We have examined sequence variation in two individuals during and after development of HIV-1-specific antibodies. Our results indicate that HIV-1 sequence variation occurs rapidly, resulting in multiple sequence variants prior to development of detectable antibodies to any viral antigens. The pattern of nucleotide sequence changes during acute infection is distinct from that observed in terminal stages of disease.

Methods

Case histories

Patient number 1 was a 29-year-old homosexual man who had been in good health until 2 weeks before

From the Department of Microbiology & Immunology, UCLA School of Medicine, *Cedars Sinai Medical Center, Los Angeles, California, †Aaron Diamond AIDS Research Center, New York, New York, and the ‡Departments of Microbiology & Immunology and Medicine, UCLA School of Medicine, Los Angeles, California, USA.

Sponsorship: This work was supported in part by a grant from the University of California Universitywide Task Force on AIDS and a Public Health Services grant from the National Institutes of Health (NIH), NS25508. I.S.Y. Chen is a Leukemia Society of American Scholar, and S. Pang is a fellow of the Fogarty International Center, NIH.

Requests for reprints to: Irvin S.Y. Chen, Division of Hematology-Oncology, Department of Medicine, UCLA School of Medicine, 10833 Le Conte Avenue, Los Angeles, CA 90024–1678, USA.

Date of receipt: 29 November 1991; revised: 24 February 1992.

© Current Science Ltd ISSN 0269-9370

admission, when he developed fever, myalgia, nausea, vomiting and diarrhea. Although the patient had had numerous sexual partners in the years preceding his presentation with symptoms, he was HIV-1-antibody-negative by enzyme immunoassay at the start of illness. He was admitted to hospital when symptom complex recurred and developed a total body rash. The patient denied sore throat, headache or neck stiffness. Temperature at the time of admission was 39.9°C, and physical examination was remarkable only for a diffuse confluent erythematous rash. On admission, serologic tests for hepatitis A and B viruses, Epstein–Barr virus, cytomegalovirus, syphilis and HIV-1 were all negative. Other laboratory tests showed 137 g/l hemoglobin and a leukocyte count of 1.5×10^9/l, with 0.63 lymphocytes and 0.24 granulocytes. The absolute CD4 lymphocyte count was 380×10^6/l, with a helper : suppressor lymphocyte ratio of 1.4. Rash and fever resolved, and leukocyte count returned to normal over the following 2 weeks.

Patient number 2 was a 28-year-old white homosexual man who had been in good health until approximately 10 days before admission, when he presented with sudden onset of fever, chills, sweats, weakness and anorexia. Four days prior to admission, the patient had been seen in an outpatient clinic and diagnosed with hepatitis. On admission, he complained of fever, nausea and four episodes of emesis. His last HIV-1-antibody test, approximately 18 months before hospital admission, was negative. The patient admitted to unprotected sexual intercourse 10 days prior to the onset of his current illness. At the time of admission, temperature was 40.2°C and he had a macular papular erythematous rash over his abdomen to the mid-axillary line. White plaques were present in oropharynx, without erythema. Neck was supple without adenopathy, and lungs were clear to auscultation and percussion. Abdominal examination revealed mild right upper quadrant tenderness, with a liver edge palpated 3 cm below the right costal margin and spanning 13 cm by percussion. Laboratory studies at the time of admission were remarkable for 140 g/l hemoglobin, 41% hematocrit, and a leukocyte count of 1.9×10^9/l, with 0.48 lymphocytes, 0.46 neutrophils, and 0.06 monocytes. Liver studies were notable for 1152 U/l aspartate aminotransferase, 590 U/l alanine aminotransferase, 174 U/l alkaline phosphatase, 56 mg/l direct bilirubin, and 6 mg/l indirect bilirubin. Serologic tests for heterophile antibodies, cytomegalovirus and Epstein–Barr virus were negative, consistent with past exposure. Hepatitis B surface antigen was negative, but immunoglobulin (Ig) M antibodies to hepatitis A were detected. Within 4 days of admission, this patient's symptoms had resolved, with gradual resolution of his rash, and return to normal levels of hepatic transaminases and bilirubin over approximately 2 weeks.

Assays for p24 antigen and HIV-1 antibodies

Sequential plasma samples from each patient were tested for p24 antigen with a commercial enzyme immunoassay (Abbott Laboratories, North Chicago, Illinois, USA), and for HIV-1-specific antibodies with a commercial enzyme immunoassay and Western blot kits (Genetic Systems, Seattle, Washington, USA).

Peripheral blood lymphocyte culture of HIV-1

Isolation of HIV-1 was performed by 4-day coculture of patient peripheral blood mononuclear cells (PBMC) with phytohemagglutinin-activated normal donor peripheral blood lymphocytes (PBL), as described previously [9].

Nucleic acid preparation

Patient blood samples were subjected to Ficoll–Hypaque density centrifugation to obtain PBL. These cells were lysed by solutions containing 0.5% sodium dodecyl sulfate (SDS), 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 0.1 mg/ml proteinase K for 3 h at 55°C. DNA was purified by two extractions each with phenol and chloroform. After ethanol precipitation, DNA was resuspended in 0.1 × TE buffer (pH 7.4). DNA concentrations were checked by measuring optical densities (OD) at 260 and 280 nm and/or agarose gel electrophoresis.

PCR

HIV-1 DNA copy number determination

HIV-1 DNA copy numbers per 100 000 cells (10^5 cells) of all DNA samples were determined by quantitative PCR [10,11], with one end-labeled (M667) and one unlabeled (M661) oligonucleotide primer pair for HIV-1, and one labeled (PC03) and unlabeled (PC04) pair for the β-globin gene [12,13]. In these studies, 0.05, 0.5, and 2.5 µg DNA isolated from PBL from HIV-1-negative blood donors served as both negative controls and cell-number references. Since each human cell contains approximately 6 pg DNA [4], 1 µg human DNA is equivalent to 1.5×10^5 cells. In our studies, a range of 0.05–2.5 µg DNA was sufficient to determine cell numbers for all samples tested. Ten to five thousand copies of cloned HIV-1 DNA, pNBJRCSF [9], served as HIV-1 copy number standards.

HIV-1 DNA amplification

After HIV-1 copy numbers per 10^5 cells were determined, 200–1000 HIV-1 copies of equivalent DNA samples were added to a PCR with unlabeled oligonucleotides [*env* 22 (GTGGAGGGGAATTCTTC-TACTGTAA) and *env* 21 (GGTACCACCATCTCTTGT-TAATAG)] [8]. The amplified region encoded one part of *env* gp120, the entire V4 region and the putative CD4-binding site (Figs 1b and 2b), corresponding to positions 7427–7580 of HIV-1_{NL4-3}. *env* 22 contains an *Eco*RI digestion site for cloning and sequencing. HIV-1 samples were amplified by PCR for 40 cycles at 94°C for 1 min, 55°C for 20 sec, and 72°C for 2 min. Uninfected PBL DNA and 500 copies of cloned HIV-$1_{JR\,CSF}$ standards were used as negative and positive controls, as described previously [11]. Since HIV-1 copy numbers were low for some samples, > 10 µg DNA was typically added to the PCR amplification. To decrease

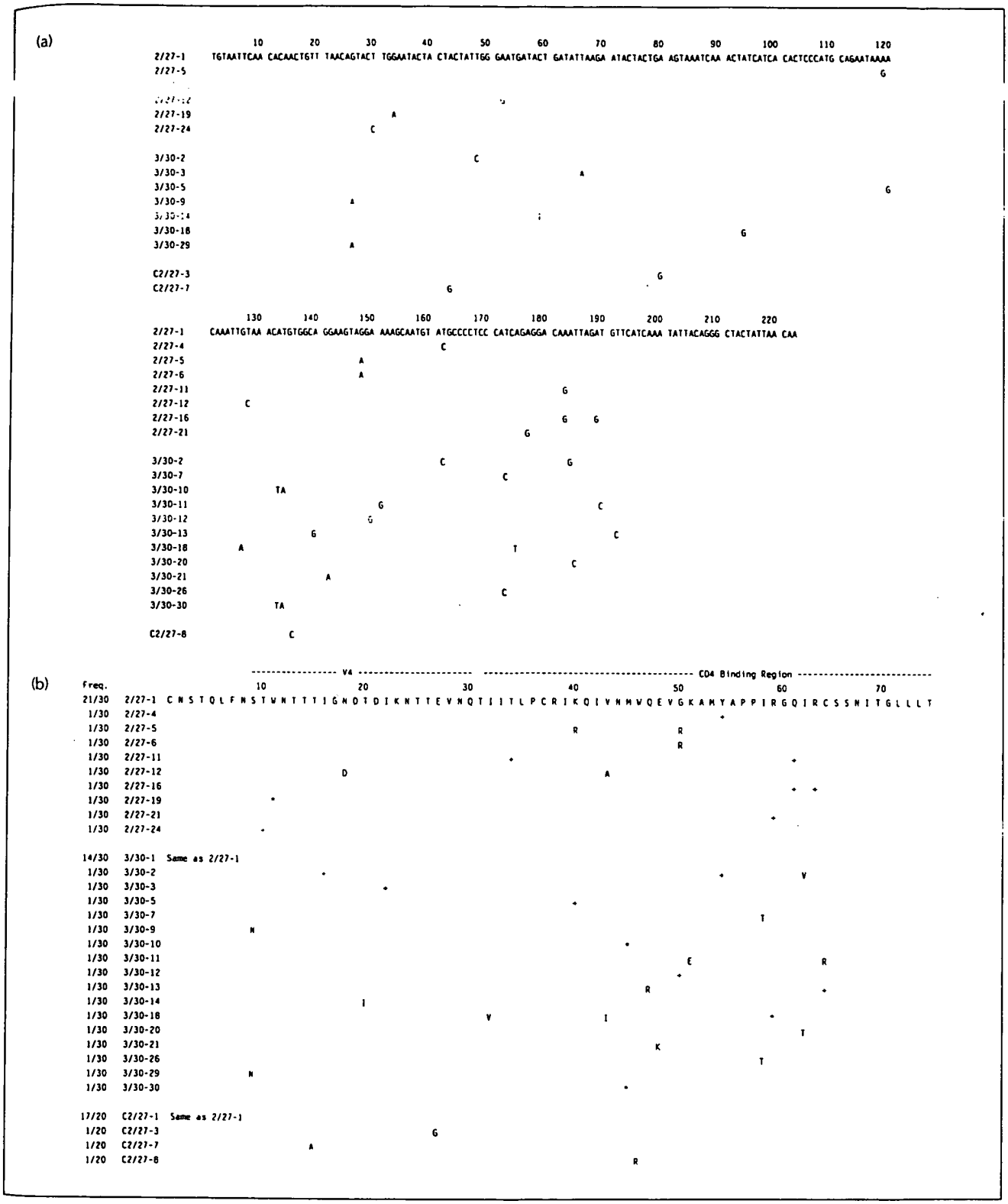


Fig. 1. Nucleic sequences and translated amino-acid sequences of HIV-1 *env* V4 and flanking regions isolated from patient number 1. Sequences are shown from the 3' of *env* 22 (CTGGAGGGGAATTCTTCTAC*TGTAA*), corresponding to position 7427 of HIV-1_{NL4-3}, and end in the middle of *env* 21 (GGTACCACCATCTC*TTGTTAATAG*), corresponding to position 7580 of HIV-1_{NL4-3}. The unique sequence isolated on 22 February 1990 is identical to 2/27-1; i.e., the dominant sequence (67%) was isolated from the sample of 27 February 1990. The dominant sequence (43%) from the sample taken on 30 March 1990 is also identical to 2/27-1. (a) Nucleic acid sequences isolated from patient 1. To shorten the sequence alignment, only the variation sites are presented; for example, sequence 2/27-12, which shows variations in two sites, A→G in position 52 and T→C in position 128. These two sites are presented in the alignment map, but sites identical to the dominant type sequence are not shown. Cultured samples passaged from 27 February 1990 HIV-1 isolates are labeled C2/27. (b) Translated amino-acid sequences isolated from patient 1. +, silent mutations.

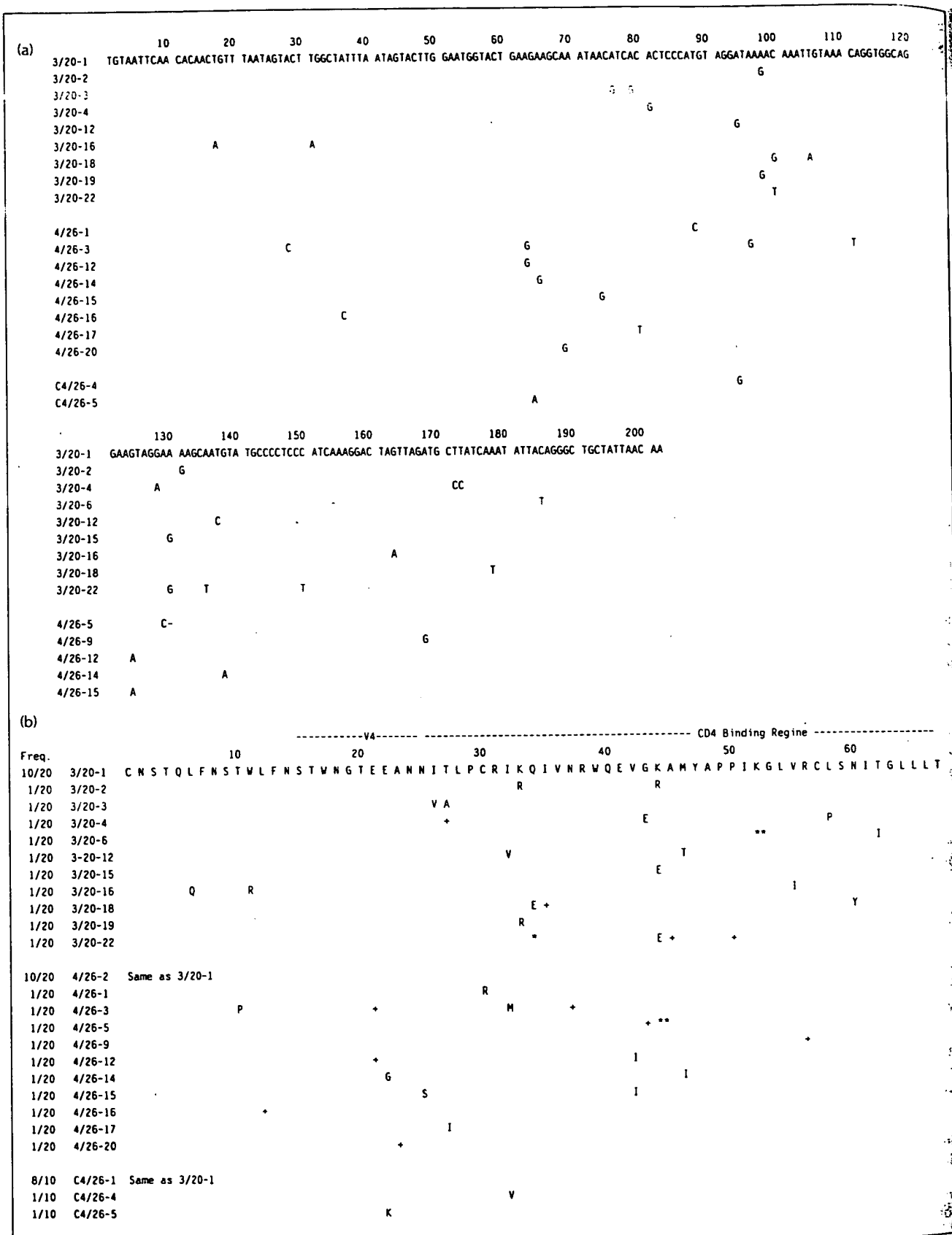


Fig. 2. Nucleic-acid sequences and translated amino-acid sequences of the HIV-1 V4 region isolated from patient number 2. Note the deletion mutations in position 154 of sequence 3,20-6 and position 130 of sequence 4/26-5 in (a). These deletions were generated with frameshifts. The corresponding sites are marked with ** in (b).

the background and increase the amplification efficiency, DNA was divided into several PCR tubes, each containing < 2.5 µg DNA.

Cloning of PCR-generated fragment

HIV-1 fragment DNA bands, generated by PCR, were demonstrated by 5% 29:1 polyacrylamide gel electrophoresis followed by ethidium bromide staining. Bands of approximately 240 base pairs (bp) were excised and eluted with STE buffer [0.1 mM NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM EDTA] at 37°C for 5–16 h. The DNA released from the gel to the buffer was precipitated by adding two volumes of ethanol and 10 mM $MgCl_2$. DNA was digested with *Eco*RI and extracted with chloroform. The resulting DNA was ligated into an M13 vector prepared by cutting M13mp19 with *Sma*I and *Eco*RI. The ligated DNA samples were transformed into *Escherichia coli* XL1-Blue-competent cells, and positive plaques screened by hybridization [8]. If the number of positive plaques was < 30, PCR was repeated.

Nucleotide sequence analysis

Sequences were analyzed using the Clustal computer program [15,16] in order to compare them with an HIV-1 database and laboratory strains. This analysis first determined lack of potential contamination, and then characterized new strains.

Results

Identification of patients with primary HIV-1 infection

Two patients were identified as possible candidates for HIV-1 infection on the basis of prolonged influenza-like illness with fever and rash in patient 1, and fever, rash and hepatitis in patient 2. Interestingly, patient 2 had concomitant acute hepatitis A and acute HIV-1 infection. These two individuals were confirmed to be infected with HIV-1 by detection of p24 antigens and identification of HIV-1 proviral DNA by PBL co-culture and PCR, respectively, prior to seroconversion (Table 1). Figure 3 illustrates quantitative PCR analysis of viral DNA levels in PBL, and Table 1 summarizes the dates at which blood was taken for assessment of viral DNA levels. For patient 1 (reported previously [17]), there was a dramatic decrease in viral DNA levels in PBL, even before development of specific antibodies to all viral proteins (Fig. 4). Relatively low levels of viral DNA were seen in each sample at each time-point assayed for patient 2.

Sequence variation of patient number 1

In a previous study, we analyzed the variable region 4 (V4) of the HIV-1 *env* gene because of its high variability relative to other variable regions of the *env* gene, and its proximity to the region involved in binding to the HIV-1 receptor, CD4 [8]. In the present study, we analyzed the same region to allow direct comparison with our previous results. DNA from each of the three time-points was subjected to PCR amplification, molecularly cloned, and individual clones isolated and subjected to nucleotide sequence analysis. In addition, a sample of viral DNA isolated following a single passage in PBL cell culture was subjected to similar sequence analysis. Thirty individual clones were sequenced from three time-points (22 February 1990, 27 February 1990, and 30 March 1990), as were 20 clones of the cultured sample (Fig. 1 and Table 2). No sequence variation was observed on 22 February 1990. Ten different species, including one major sequence and nine variants, were observed out of 30 clones on 27 February 1990, and 17 out of 30 clones were observed on 30 March 1990. In the cultured sample, four different sequences out of 20 were observed. The major species was identical in the cultured sample throughout, and each of the individual minor sequences was represented once. A few frameshift mutations resulting in premature termination were also observed, presumably representing defective proviruses.

It is particularly noteworthy that the variation observed on 27 February 1990 was evident only 5 days after 22 February 1990, when no variations were observed.

Table 1. Summary of virological data in primary infection.

Sample	Days since onset of illness	HIV-1 status	p24 (pg/ml)	Plasma (TCID/ml)	Cells (TCID/10^6cells)	HIV-1 DNA level (copies/10^5cells)
Patient 1						
PBL	5 (2/17/90)	−	4200	NT	10^4	NT
PBL	10 (2/22/90)	−	1450	1000	10^4	3000
PBL	15 (2/27/90)	−	2069	100	10^4	60
PBL	47 (3/30/90)	+	< 50	1	10	20
Culture	15 (2/27/90)					100
Patient 2						
PBL	12 (3/20/91)	−	5312	< 500	> 5000	
PBL	49 (4/26/91)	+	< 50	100	500	
Culture	49 (4/26/91)					

The number of HIV-1 copies per 10^5 cells was determined by quantitative polymerase chain reaction, as described in Methods. The cultured samples were derived by 4 days of cocultivation of peripheral blood lymphocytes (PBL) obtained at the indicated dates from two patients with PBL from uninfected donors. For patient 1, p24, plasma and cell tissue culture infective dose (TCID) data have been published [17]. NT, not tested. − negative; + positive.

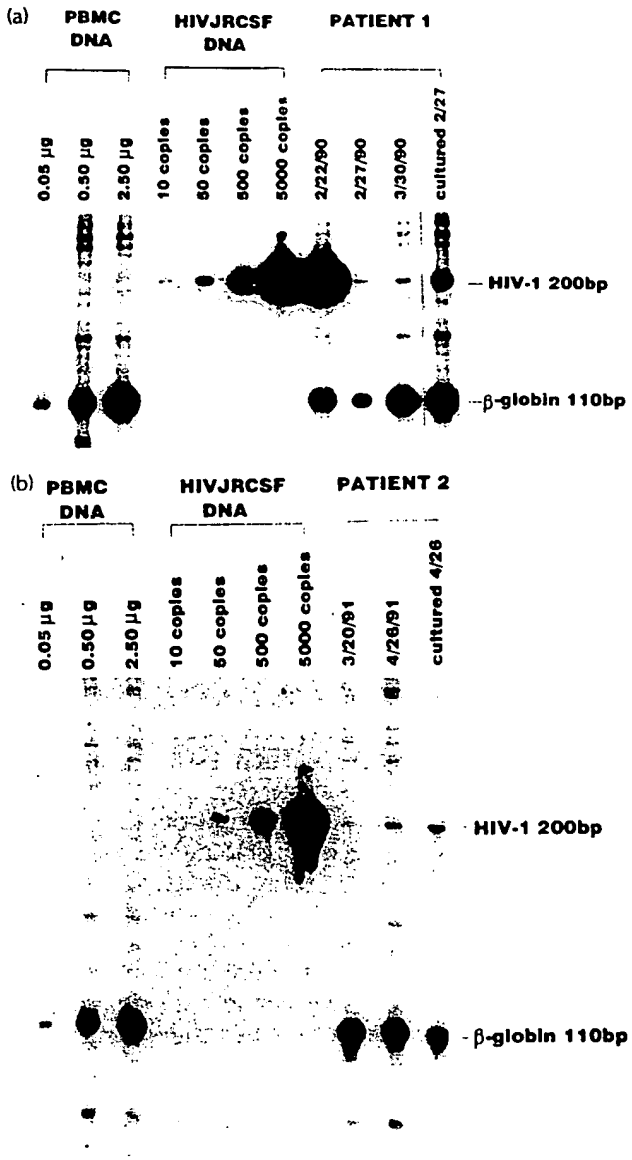


Fig. 3. (a) HIV-1 DNA from patient number 1 blood and culture samples. Each polymerase chain reaction (PCR) contained 5×10^6 c.p.m. ^{32}P-end-labeled M667 to oligonucleotide (approximately 30 ng), 10^6 c.p.m. ^{32}P-endlabeled PC03, 100 ng M661, and 20 ng PC04. Peripheral blood lymphocyte (PBL) DNA isolated from an uninfected blood donor served as both negative control and cell-number standard. Cloned HIV-1, pNB-JRCSF, served as both positive control and HIV-1 copy number standard. PCR buffer contained 50 mM NaCl, 5 mm $MgCl_2$, 250 µM dNTP, and 25 mM Tris (pH 8.0). Amplifications were performed at 94°C for 1 min and 65°C for 2 min for 25 cycles. (b) HIV-1 DNA from patient number 2 blood and from cultured samples. The assay conditions were the same as in (a).

Thus, the emergence of variants in a significant proportion of the HIV-1 sequences occurred very rapidly. Given the major decrease in viral DNA levels from the time-points 22 to 27 February 1990 (Fig. 3 and Table 1), it is possible that variants were pre-existing but detected only after the major species was reduced in level; however, we cannot distinguish this from the possibility that the variants emerged *de novo* within 5 days.

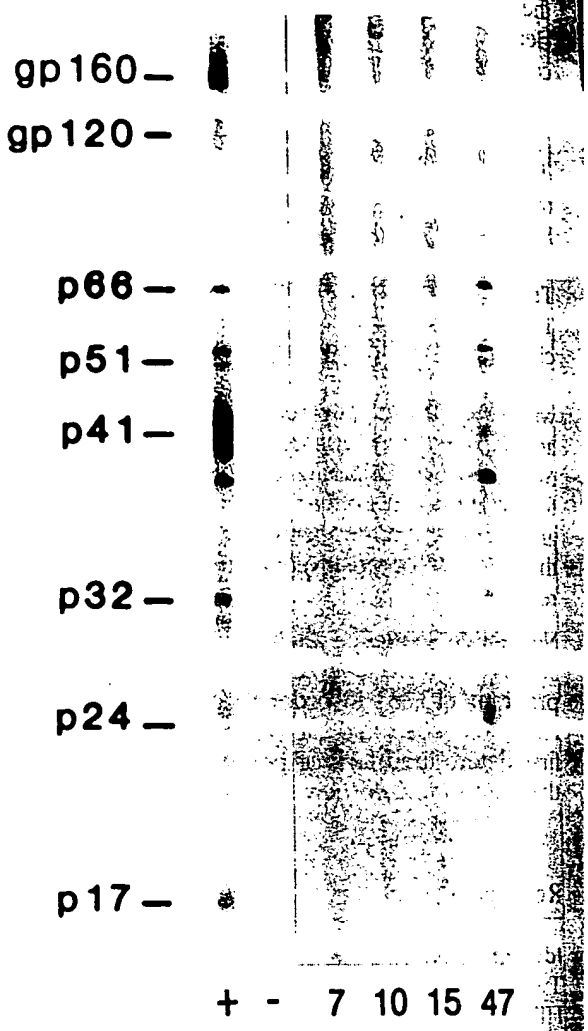


Fig. 4. Serial Western blots demonstrating HIV-1 seroconversion in patient number 1. The numbers at the bottom of the figure represent the day of illness on which the serum sample was collected. +, HIV-1-positive human serum; −, normal human serum.

The last time-point was subsequent to development of antibodies to major HIV-1 proteins as detected by Western blot. We observed no significant differences in patterns of sequence variation between the samples obtained before and after seroconversion.

Sequence variation of patient number 2

For the second individual, two time-points and a cultured sample were subjected to nucleotide sequence analysis. The first time-point was prior to detection of HIV-1-specific antibodies, and the second subsequent to seroconversion. Plasma titers, measured by limiting dilution, decreased from 500 to 100 IU/ml between the two time-points, and a proportion of infected cells decreased from 5×10^3 cells to 5×10^2 IU/10^6 cells. Twenty clones were sequenced from each time-point, resulting in 10 out of 20 variants at the first time-point, and 10 out of 20 at the second. Thus, as for patient 1, significant sequence variation was observed during the acute phase of infection.

Table 2. Summary of HIV-1 genetic variation in V4/CD4 binding regions.

Sample	Days since onset of illness	HIV-1 variants (no. total NT sequences)	HIV-1 variants (no. total AA sequences)	Silent mutations (mean mutations)
Patient 1				
PBL	10 (2/22/90)	0/30	0/30	0
PBL	15 (2/27/90)	9/30	4/30	7.13
PBL	47 (3/30/90)	16/30	13/30	6.21
Culture	15 (2/27/90)	3/20	3/20	
Patient 2				
PBL	12 (3/20/90)	10/20	10/20	4.21
PBL	49 (4/26/90)	10/20	8/20	7.16
Culture	49 (4/26/90)	2/10	2/10	0.2

HIV-1 copy number per 10^5 cells. Cloned HIV-1 DNA were derived from preparative polymerase chain reaction (PCR) as described in Methods. The major sequence of patients 1 and 2 from all samples was identical, and considered to be the node sequence. NT, nucleotide; AA, amino acid.

Frequency of silent mutations

Several groups have shown that when HIV-1 sequence analysis is performed on sequences in blood or brain of AIDS patients, the proportion of silent mutations is much lower than would be expected based upon random occurrence. In three previous individuals we studied, for example, only 1–5% of the mutations in V4 were silent [8]. Similar observations for the V3 and V4–V5 regions have been reported by other groups [18,19]. Among the variants described here, the proportion of silent mutations was much higher than that seen previously, and was consistent with what might be expected from random occurrence (≈25%) [20]. In previous studies, the low frequency of silent mutations suggested that selection pressures for amino-acid changes were playing a role in the emergence of particular HIV-1 variants [8,18–21]. Here, the increased incidence of silent mutations indicates the converse, that in these early cases of HIV-1 infection, selection, at least as measured in the peripheral circulation, is not yet playing a major role in emergence of variants within V4 and flanking regions. It is also notable that the G to A and C to T hypermutation rates are much lower than in previous reports [1,7,21]. The G to A mutation frequency was 8 out of 40 for patient 1, and 11 out of 43 for patient 2. A higher frequency of A to G and T to C mutations was observed (25 out of 40 for patient 1; 21 out of 43 for patient 2).

Discussion

Our results indicate that following HIV-1 infection, a single viral strain predominates, either because it was the major strain in the infecting inoculum or because of its greater replicative potential. The precise time of infection was unknown for the two individuals studied, but is likely to have been no more than a few months prior to seroconversion. The sequence variation observed therefore occurred within a few months after infection. It appears likely that the variation observed is due to mutation occurring *de novo*, since the emergence of variants appears to be an extremely dynamic process. Even within the relatively short time frame of this analysis, none of the minor variants observed were conserved from one time-point to the next, whereas the major species of HIV-1 was maintained. However, since we do not have direct information concerning the pattern of variation in the donor, we cannot exclude the formal possibility that all of the variants were from the infecting inoculum.

It is of interest to compare the pattern of variation observed in these cases of acute HIV-1 infection with that observed in symptomatic patients infected months to years prior to analysis. First, each variant is represented only once in our analysis, and variants cannot easily be divided into related subgroups by computer analysis, as was the case in our previous analysis [8]. This probably reflects the relatively early evolution of variants from a common progenitor in the case of acute infection, whereas months to years after infection, low-level replication with time results in a more 'branched' evolution. Second, we observed a much higher incidence of silent mutations compared with advanced HIV-1 infection. The frequency of silent mutations often indicates whether selection processes are playing a role in evolution [18,19,21–23]. It would appear that in the early stages of HIV-1 infection, at least within the V4 region, there is no apparent selection for amino-acid changes.

Variation was observed in both the individuals studied before detection of antibodies to viral proteins by Western blot. This suggests that the humoral immune response does not play a direct role in the selection of particular variants. Thus, mechanisms for rapid emergence and disappearance of individual variants are unclear. The decline in viral load may be due to the response to cytotoxic T-lymphocytes [24,25] or natural killer cells, to limited activated CD4+ T-cells permissive for infection [26], or to mechanisms that remain to be explained.

Acknowledgements

We would like to thank Dr W. O'Brien, Dr T. Zhu, Dr D. Camerini, and Dr J. Zack for helpful suggestions; A. Haislip and Y. Xie for technical assistance; and W. Aft for manuscript preparation.

References

1. Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S: HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J Acquir Immune Defic Syndr* 1989, 2:344–352.
2. Saag MS, Hahn BH, Gibbons J, *et al.*: Extensive variation of human immunodeficiency virus type 1 *in vivo*. *Nature* 1988, 334:440–444.

3. ALIZON M, WAIN-HOBSON S, MONTAGNIER L, SONIGO P: Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 1986, 46:63–74.
4. STARCICH BR, HAHN BH, SHAW GM, *ET AL*: Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 1986, 45:637–648.
5. WILLEY RL, RUTLEDGE RA, DIAS S, *ET AL*: Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. *Proc Natl Acad Sci USA* 1986, 83:5038–5042.
6. HAHN BH, GONDA MA, SHAW GM, *ET AL*: Genomic diversity of the acquired immune deficiency syndrome virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. *Proc Natl Acad Sci USA* 1985, 82:4813–4817.
7. MEYERHANS A, CHEYNIER R, ALBERT J, *ET AL*: Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. *Cell* 1989, 58:901–910.
8. PANG S, VINTERS HV, AKASHI T, O'BRIEN WA, CHEN ISY: HIV-1 Env sequence variation in brain tissue of patients with AIDS encephalopathy. *J Acquir Immune Defic Syndr* 1991, 4:1082–1092.
9. KOYANAGI Y, MILES S, MITSUYASU RT, MERRILL JE, VINTERS HV, CHEN ISY: Dual infection of the central nervous system by AIDS viruses with distinct cellular tropisms. *Science* 1987, 236:819–822.
10. LEE H, SWANSON P, SHORTY VS, ZACK JA, ROSENBLATT JD, CHEN ISY: High rate of HTLV-II infection in seropositive IV drug abusers from New Orleans. *Science* 1989, 244:471–475.
11. PANG S, KOYANAGI Y, MILES S, WILEY C, VINTERS H, CHEN ISY: The structure of HIV DNA in brain and blood of AIDS patients. *Nature* 1990, 343:85–90.
12. SAIKI RK, SCHARF S, FALOONA F, *ET AL*: Enzymatic amplification of β-globin genomic sequences and restriction site analysis of diagnosis of sickle-cell anemia. *Science* 1985, 230:1350–1354.
13. SAIKI RK, GELFAND DH, STOFFEL S, *ET AL*: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988, 239:487–491.
14. LEHNINGER AL: *Biochemistry*: 2nd edn., edited by Lehninger AL. New York: Worth Publishers Inc., 1975.
15. HIGGINS DG, SHARP PM: Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene* 1988, 73:237–244.
16. HIGGINS DG, SHARP PM: Fast and sensitive multip alignments on a microcomputer. *CABIOS* 1989.
17. DAAR ES, MOUDGIL T, MEYER RD, HO DD: Tran levels of HIV-1 in the blood of patients with fection. *N Engl J Med* 1991, 324:961–964.
18. SIMMONDS P, BALFE P, LUDLAM CA, BISHOP JO, BROW ysis of sequence diversity of hypervariable reg external glycoprotein of human immunodefic type 1. *J Virol* 1990, 64:5840–5850.
19. WOLFS TFW, ZWART G, BAKKER M, VALK M, GOUDSMIT J: Naturally occurring mutations withi genomic RNA lead to antigenic variation depe single amino-acid substitution. *Virology* 1991, 18
20. JUKES TH, KING JL: Evolutionary nucleotide repla DNA. *Nature* 1979, 281:605–606.
21. MYERS G, PAVLAKIS GN: Evolutionary potential o retroviruses. In *Viruses* edited by Wagner RR, Fr rat H. New York: Plenum Press, 1991.
22. BROWN AL, MONAGHAN P: Evolution of the structur of human immunodeficiency virus: selective con nucleotide substitution. *AIDS Res Hum Retrovi* 4:399–407.
23. BALFE P, SIMMONDS P, LUDLAM CA, BISHOP JO, BROW current evolution of human immunodeficiency vi in patients infected from the same source: rate of change and low frequency of inactivating mutati 1990. 64:6221–6233.
24. LITTAUA RA, OLDSTONE MB, TAKEDA A, ENNIS FA: cytotoxic T-lymphocyte clone to a conserved e human immunodeficiency virus type 1 p24: cytot ity and secretion of interleukin-2 and interleukin 1992. 66:606–611.
25. LITTAUA RA, OLDSTONE MB, TAKEDA A, *ET AL*: An stricted CD8+ cytotoxic T-lymphocyte clone r a highly conserved epitope on human immunod virus type 1 *gag*. *J Virol* 1991, 65:4051–4056.
26. ZACK JA, ARRIGO SJ, WEITSMAN SR, GO AS, HAISLIP A, HIV-1 entry into quiescent primary lymphocytes: analysis reveals a labile, latent viral structure. 61:213–222.

SN 643,306
filed 8/22/84

APPLICATION FOR UNITED STATES PATENT

Inventors: FLOSSIE WONG-STAAL
ROBERT C. GALLO
BEATRICE H. HAHN
Mikulas Popovic

Title: MOLECULAR CLONES OF THE GENOME OF HTLV-III

Abstract of the Disclosure

Disclosed is the molecular cloning of HTLV-III, the adult leukemia and acquired immune deficiency syndrome (AIDS) virus. Clone BH10 contains a 9.0 Kb viral insert constituting the entire HTLV-III genome. Clones BH8 and BH5 contain viral inserts of 5.5 Kb and 3.5 Kb, respectively. These clones are suitable for the development of diagnostic and therapeutic measures for AIDS, as well as use as probes for the detection of AIDS.

In related inventions, HTLV-III was detected, isolated, and immortalized in an HT cell line. Since evidence now strongly indicates that HTLV-III is related to acquired immune deficiency syndrome (AIDS), the ability to enhance production of the virus and determine the DNA sequences of the virus is critically important to developing a cure or reagent active against AIDS. The present invention takes one such significant step by disclosing the process for molecularly cloning the complete genome of the HTLV-III virus. In short, The molecular cloning of the complete genome of the HTLV-III virus produced by one of these lines designated H9/HTLV-III is disclosed. Two forms of this virus are identified which are highly related but differ in several restriction enzyme cleavage sites. Both variants exist as integrated and unintegrated forms in the infected cell line. The complete genomes of two forms of HTLV-III are molecularly cloned and shown to exist in the long-term infected cell line both as polyclonally integrated provirus and as unintegrated viral DNA. These clones are used as probes to detect viral sequences in cell lines other than H9/HTLV-III, taken from different AIDS patients, and in fresh lymphoid tissues of AIDS patients, providing further evidence that the cloned genomes constitute predominant forms of HTLV-III, the causative agent in AIDS.

Statement of Utility

Previous work with the HTLV family of virus showed three variants. Of these, it was believed that HTLV-III was the causative agent of AIDS. Using the clones produced by this invention, HTLV-III has been shown to be distinctly different than HTLV-I and HTLV-II, whereas HTLV-I and -II share greater homology and thus better identification of AIDS virus in sera.

Description of the Figures

Figure 1 is a Southern blot analysis of unintegrated DNA of HTLV-III. No viral sequences could be detected in the undigested DNA after 4 hours. However, a major species of viral DNA of approximately 10 Kb length was present in the 10, 15, 24 and 48 hr harvest representing the linear unintegrated form of the virus. A representative Southern blot of the 15 hr harvest digested with several restriction enzymes is shown in this figure. Methods: 8 x 10^8 fresh uninfected H9 cells were infected with concentrated supernatant from cell line H9/HTLV-III containing 4 X 10^{11} particles of HTLV-III. Infected cells were divided into five Roller bottles and harvested after 4, 10, 15, 24 and 48 hrs. Low molecular weight DNA was prepared using the Hirt fractionation procedure and 30 ug of undigested and digested DNA were separated on a 0.8% agarose gel, transferred to nitrocellulose paper and hybridized to a HTLV-III cDNA probe for 24 hr at 37°C in 1 X SSC, 40% formamide and 10% Dextran sulfate. cDNA was synthesized from poly(A) selected RNA prepared from doubly banded HTLV-III virus in the presence of oligo(dT) primers. Filters were washed at 1 X SSC at 65°C.

Figure 2 is a restriction endonuclease map of two closely related HTLV-III variants cloned from unintegrated viral DNA. Three recombinant clones (λ BH10, λBH5 and λBH8) were analyzed and their inserts (9 Kb, 5.5 Kb and 3.5 Kb, respectively) were mapped with the indicated enzymes. They represent two variant forms of HTLV-III differing in three enzyme sites which are depicted in bold letters and by an asterisk. As SstI cuts the LTR of the HTLV-III the three clones represent two full length genomes with one LTR. A schematic map of this viral genome is shown at the bottom of the

figure, although the total length of the LTR is approximate. Methods: Low molecular weight DNA combined from the 15 and 24 hr harvest was fractionated on a 10-40% sucrose gradient. Aliquots of the fractions were electrophoresed on a 0.5% agarose gel, transferred to nitrocellulose paper and hybridized to HTLV cDNA under conditions described in Figure 1. Fractions which contained the unintegrated linear HTLV-III genome shown by hybridization were pooled, the DNA was subsequently digested with SstI and ligated to phosphatase treated SstI arms of λgtWes·λB. After in vitro packaging, recombinant phages were screened for viral sequences with HTLV-III cDNA.

Figure 3 demonstrates HTLV-III viral sequences in the infected cell line H9/HTLV-III. Both variant forms of HTLV-III were detected as integrated provirus as well as unintegrated viral DNA in the infected cell line. However, no viral sequences were found in uninfected H9 cells, uninfected HT cells nor in normal human thymus (NT). Methods: 10 µg of high molecular weight DNA were digested with restriction enzymes as indicated and hybridized to nick translated phage insert from BH10 under the same conditions as described in Figure 1.

Figure 4 shows a sequence homology of HTLV-III to other members of the HTLV family. A schematic restriction map of HTLV-I, HTLV-Ib and HTLV-II is drawn below indicating the length and the location of the generated fragments in respect to the corresponding genomic regions. LTR, gag, pol, env and pX regions are drawn to scale according to the published nucleotide sequence of HTLV-I. The bands which are most highly conserved as stringency increases correspond to the gag/pol junction region of HTLV-I (1.8 Kb PstI fragment and HTLV-IIb (3.1 Kb PstI fragment) and to the 3' part

of the pol region of HTLV-II (2.1 Kb SmaI/BamHI fragment) which do not overlap assuming the same genomic organization in HTLV-II. Fragments corresponding to pX of HTLV-I (2.1 Kb SstI Pst fragment) and HTLV-Ib (1.4 Kb Pst fragment) are less conserved but still visible at Tm - 28°C on the original autoradiogram. Digestion of GaLV generates six fragments, none of which show hybridization under medium or high stringency. Methods: Subclones of full length genomes of a prototype HTLV-I, HTLV-Ib, HTLV-III and GaLV (Seato strain) were digested with the following enzymes, PstI plus SstI (HTLV-I and HTLV-Ib), BamHI plus SmaI (HTLV-II) and Hind III plus SmaI plus XhoI (GaLV). Four replicate filters were prepared and hybridized for 36 hr under low stringency (8 X SSC, 20% formamide, 10% Dextran sulfate at 37°C) to nick translated insert of λ BH10. Filters were then washed in 1 X SSC at different temperatures, 22°C (Tm - 70°C) filter 1, 37°C (Tm - 56°C) filter 2, 50°C (Tm - 42°C) filter 3 and 65°C (Tm - 28°C).

The Invention

The present invention discloses a method for production of molecular clones of HTLV-III from a fraction enriched for the unintegrated provirus in acutely infected cells. Three clones for the HTLV-III genome were produced using recombinant DNA techniques by isolating and characterizing unintegrated viral DNA, cleaving this DNA with the appropriate restriction enzyme, and constructing a phage library capable of being screened by viral cDNA. This process led to the production of three clones: BH10, containing a viral insert of 9.0 Kb corresponding to the complete HTLV-III genome; clone BH8 containing an insert of 5.5 Kb; and clone BH5 containing a viral insert of 3.5 Kb. See Figure 3 for a pictoral representation of the differences between these three clones.

In general, cloning the HTLV-III genome involved isolating unintegrated viral DNA after infection of H9-cells with concentrated HTLV-III virus and cloning this DNA in a lambda phage library to be screened with viral cDNA. The cell line H9/HTLV-III produces large quantities of HTLV-III virus and serves as the principal producer cell line for immunological assays used to detect virus specific antigens and antibodies in AIDS sera. Cultures of H9/HTLV-III cells (infected cells) are grown and harvested, followed by extraction of low molecular weight DNA from the newly infected cells. This produces unintegrated viral DNA. A cDNA library is formed using HTLV-III cDNA. This cDNA is then used as a probe for assaying the unintegrated viral DNA. Unintegrated linear DNA (provirus DNA) is then obtained, containing the entire HTLV-III genome, i.e., replication competent. This DNA is then digested in plasmid lambda gt Wes · lambda B to form clone lambda BH10. The other clones are produced by digesting provirus DNA that does not contain the entire HTLV-III genome.

Two elements of the above process are recombinant DNA procedures, such as, the DNA library and a cDNA probe. The library is formed by taking the total DNA from H9/HTLV-III cells, cutting the DNA into fragments with a suitable restriction enzyme, hybridizing to the fragments to a radiolabeled cDNA probe, joining the fragments to plasmid vectors, and then introducing the recombinant DNA into a suitable host.

The cDNA probe is an HTLV-III cDNA probe made from double-banded HTLV-III mRNA. A short oligo-dT chain is hybridized to the poly-A tail of the mRNA strand. The oligo-T segment serves as a primer for the action of reverse transcriptase, which uses the mRNA as a template for the synthesis of a complementary DNA strand. The resulting cDNA ends in a hairpin loop.

Once the mRNA strand is degraded by treatment with NaOH, the hairpin loop becomes a primer for DNA polymerase I, which completes the paired DNA strand. The loop is then cleaved by S1 nuclease to produce a double-stranded cDNA molecule. Linkers are then added to the double-stranded cDNA by using DNA ligase. After the linkers are cut open with a restriction enzyme and the cDNA is inserted into a suitable plasmid cleaved with the same enzyme, such as pBR322. The result is a cDNA-containing recombinant plasmid.

Statement of Deposit

The cell lines and clones of this invention are on deposit in the American Type Culture Collection in the manner prescribed by the Patent and Trademark Office with regard to permanence of the deposit for the life of the patent and without restriction on public access. The accession numbers are: H9/HTLV-III, CRL 8543; BH10, #40125; BH8, #40127; and BH5, #40126.

Specific Disclosure

Concentrated virus from H9/HTLV-III is used to infect fresh uninfected H9 cells at a multiplicity of 50 viral particles/cell; cultures are harvested after 4, 10, 15, 24 and 48 hours. Extrachromosomal DNA is extracted according to the procedure of Hirt and assayed for its content of unintegrated viral DNA using HTLV-III cDNA as a probe. This cDNA is primed by oligo(dT) and copied from poly(A) containing RNA from virions that had been twice banded on sucrose density gradients. Unintegrated linear viral DNA is first detected after 10 hrs and is also present at the subsequent time points. A Southern blot of the 15 hr harvest is shown in Figure 1. A band of approximately 10 Kb in the undigested DNA represents the linear form of the unintegrated,

replication-competent HTLV-III. No closed or nicked circular DNA could be detected in the 10, 15 and 24 hour harvest, but both forms were evident in small amounts at the 48 hr harvest (data not shown). The viral genome was not cut by XbaI, whereas SstI generated three predominant bands of 9 Kb, 5.5 Kb and 3.5 Kb (Figure 1). These bands represent the complete genomes of two forms of HTLV-III, both cut by SstI in the LTR and one having an additional SstI site in the middle of its genome. Clone BH10 contains a viral insert of 9.0 Kb, a size consistent with the complete HTLV-III genome. Clones BH8 and BH5 contain inserts of 5.5 Kb and 3.5 Kb, respectively, and together they overlap completely with BH10, except for a few restriction enzyme sites polymorphisms in BH5. Therefore, BH10 and BH8 plus BH5 represent two variants of HTLV-III.

EXAMPLE 1

In order to demonstrate the presence of these two variants in the original cell line, nick-translated inserts of lambda BH10 was hybridized to a Southern blot of H9/HTLV-III genomic cDNA digested with several restriction enzymes (Figure 3). Both forms could be detected using the enzyme SstI generating the expected 3 bands of 9.0 Kb, 5.5 Kb and 3.5 Kb Xba which does not cut the provirus generating a high molecular weight genome representing polyclonal integration of the provirus and a band of approximately 10 Kb which could be interpreted as representing unintegrated viral DNA since a band of identical size was also present in the undigested first preparation (Figure 1). This was confirmed by Southern blot hybridization of undigested cellular DNA. The existence of unintegrated viral DNA thus explains the presence of a 4 Kb and 4.5 Kb EcoRI fragment seen in both first and total cellular DNA preparations (Figure 1 and Figure 3). BglII and HindIII

both cut the LTR and generated the expected internal bands. Several faint bands in the HindIII digest, in addition to the internal bands, represent either defective proviruses or another variant form with differences in the HindIII restriction pattern. The lack of HTLV-III sequences in the uninfected H9 cell line and the uninfected parental line HT as well as in normal human thymus demonstrated the exogenous nature of HTLV-III and showed that the virus does not contain any human cellular sequences. The same results were obtained using nick-translated inserts from lambda BH5 and lambda BH8.

EXAMPLE 2

The availability of the cloned HTLV-III genome allowed sequence homology between HTLV-III, HTLV-I and HTLV-II to be evaluated. Replicate Southern blots of restriction enzyme digested clones representing the complete genomes of HTLV-I, HTLV-Ib, HTLV-II and GALV as a control were hybridized to full length HTLV-III probe under relaxed conditions. The filters were then washed (λ BH10i) under conditions of low, medium, and high stringencies in order to estimate the extent of homology between HTLV-III and these viruses (Figure 4). This experiment showed that there is specific homology between HTLV-III, HTLV-I, HTLV-Ib and HTLV-II but not with HTLV-III and GALV. As demonstrated, hybridization of HTLV-III to other members of the HTLV family could be detected at the values of -42°C and -28°C, conditions under which no hybridization to GALV was seen (Figure 4, panels C and D). Of note, the restriction fragments showing greatest homology correspond to the gag/pol region of HTLV-I and to an apparently non-overlapping portion of the pol region of HTLV-II (assuming that the genomic arrangement is similar to that of HTLV-I). Further analysis revealed that it is the 5' half of the

gag and the gap between gag and pol which has the greatest homology in HTLV-I. Finally, in HTLV-Ib (a variant of HTLV-I) hybridization to the px region could be seen (1.4 Kb Pst fragment) as well as to the corresponding px fragment in HTLV-I (2.1 Kb Pst/Sst) on the original autoradiogram.

EXAMPLE 3

Figure 2 shows the restriction map of three clones designated λBH10, λBH5 and λBH8 which correspond in size to the three SstI fragments shown in Figure 1. Comparison of these maps suggests that λBH5 plus λBH8 constitute one HTLV-III genome, and λBH10 another. The two viral forms differ in only three out of 21 mapped enzyme sites, including the internal SstI site. As expected, the phage inserts of λBH5 and λBH8 hybridize under high stringency conditions to λBH10 but not to each other as analyzed by Southern blot hybridization and electron microscopic heteroduplex analysis. To show the presence of LTR sequences in the clones and to determine their orientation, a cDNA clone (C15) was used as a probe and contained U3 and R sequences. This clone strongly hybridized to the 0.5 Kb BglII fragment of λBH10 and λBH8, orienting this side 3', and faintly hybridized to the 0.7 Kb SstI/PstI fragment of λBH5 and λBH10, orienting this side 5', and demonstrated that SstI cuts the LTR of HTLV-III in the R region.

EXAMPLE 4

The presence of two variant forms of HTLV-III in the original cell line was demonstrated by hybridizing the radiolabelled insert of λBH10 to a Southern blot of H9/HTLV-III genomic DNA digested with several restriction enzymes (Figure 3). Both forms were detected using the enzyme SstI which generated the

expected 3 bands of 9Kb, 5.5 Kb and 3.5 Kb length. Both of these forms are also present as integrated proviruses because they have been cloned along with their flanking cellular sequences from a genomic library of H9/HTLV-III. Furthermore, XbaI, which does not cut the provirus, generated a high molecular weight smear representing polyclonal integration of the provirus and a band of approximately 10 Kb, representing unintegrated viral DNA. This same 10 Kb band was also detected in undigested H9/HTLV-III DNA, again indicating unintegrated viral DNA. The presence of unintegrated viral DNA also explains the 4 Kb and 4.5 Kb EcoRI fragment seen in both the Hirt and total cellular DNA preparations (Figures 1, 3). Bgl II and Hind III both cut the LTR and generate the expected internal bands. Several faint bands, in addition to the internal bands using Hind III, represent either defective proviruses or another variant form present in low copy number. The lack of HTLV-III sequences in the DNA of the uninfected H9 cell line and the uninfected parental cell line HT as well as in normal human thymus clearly demonstrates the exogenous nature of HTLV-III and shows that the virus does not contain human cellular sequences. The same results were obtained using λBH5 and λBH8 as probe inserts.

WE CLAIM

1. Recombinant clone BH10 characterized by containing the complete HTLV-III genome.

2. Recombinant clone BH8 characterized by containing a 5.5 Kb viral insert from HTLV-III virus.

3. Recombinant clone BH5 characterized by containing a 3.5 Kb viral insert from HTLV-III.

4. A process for the production of recombinant molecular clones of HTLV-III consisting essentially of cleaving unintegrated viral DNA from HTLV-III cells with a restriction enzyme to obtain a provirus, hybridizing radiolabeled cDNA to said provirus, and digesting said virus in a suitable plasmid.

5. A process for the molecular cloning and expression of a cDNA sequence of HTLV-III consisting essentially of

isolating total cellular mRNA from H9/HTLV-III cells;

forming double-stranded cDNA from said mRNA and inserting said double-stranded cDNA into a phage lambda to form a recombinant DNA molecule;

hybridizing said recombinant DNA molecule with a radiolabelled probe;

removing cDNA from said molecules and inserting said cDNA into a suitable plasmid; and

transfecting said plasmids into a suitable host cell capable of expressing HTLV-III DNA sequences.

6. A process of Claim 5 wherein said plasmid is λBH10.

7. A process of Claim 5 wherein said plasmid is λ BH8.

8. A process of Claim 5 wherein said plasmid is λ BH5.

9. A process of Claim 5 wherein said cDNA sequence corresponds to a 9.0 Kb sequence.

10. A process of Claim 5 wherein said cDNA sequence corresponds to a 5.5 Kb sequence.

11. A process of Claim 5 wherein said cDNA sequence corresponds to a 3.5 Kb sequence.

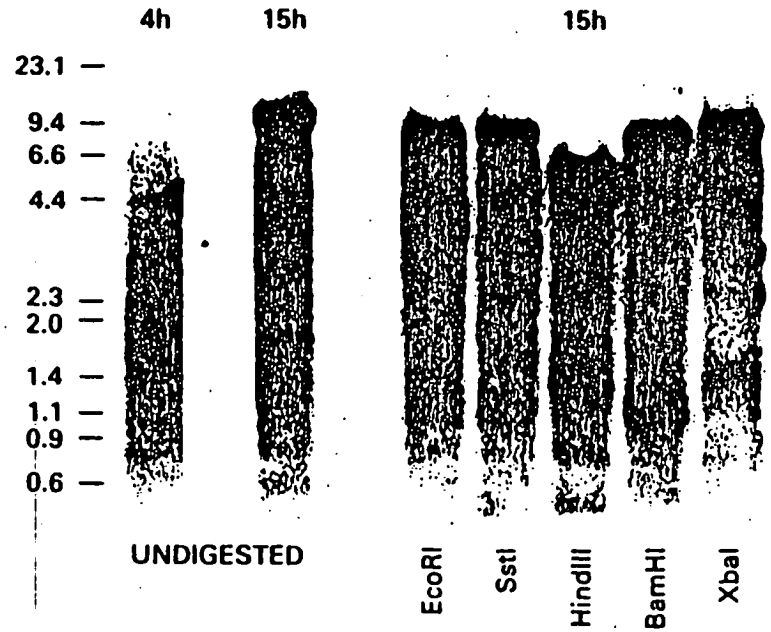

4h
15h
15h
23.1
9.4
6.6
4.4
2.3
2.0
1.4
1.1
0.9
0.6
UNDIGESTED
EcoRI
SstI
HindIII
BamHI
XbaI

FIG 1

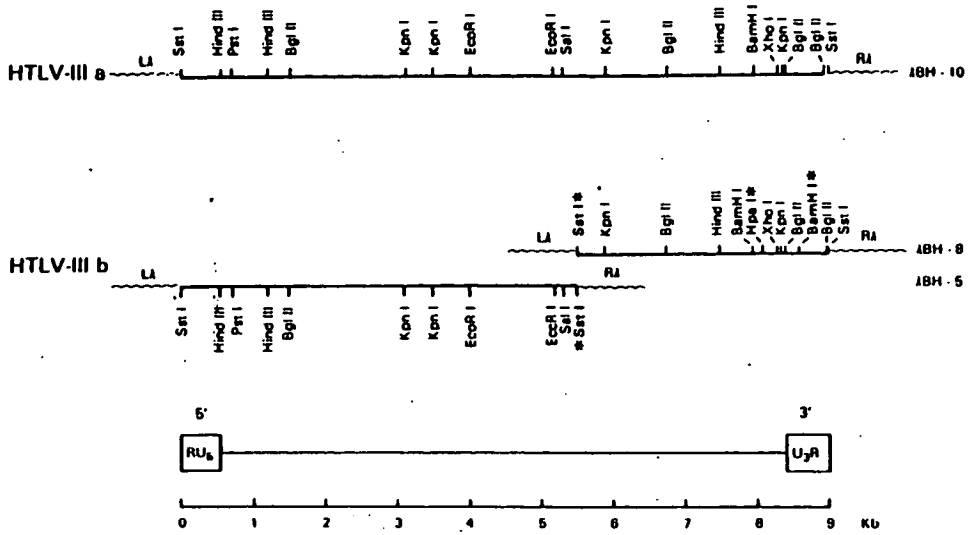

HTLV-III a
HTLV-III b
LI
RI
λBH - 10
λBH - 8
λBH - 5
Sst I
Hind III
Pst I
Bgl II
Kpn I
EcoR I
Sal I
BamH I
Hpa I
Xho I
5'
3'
RU5
U3R
0
1
2
3
4
5
6
7
8
9
Kb

Fig 7

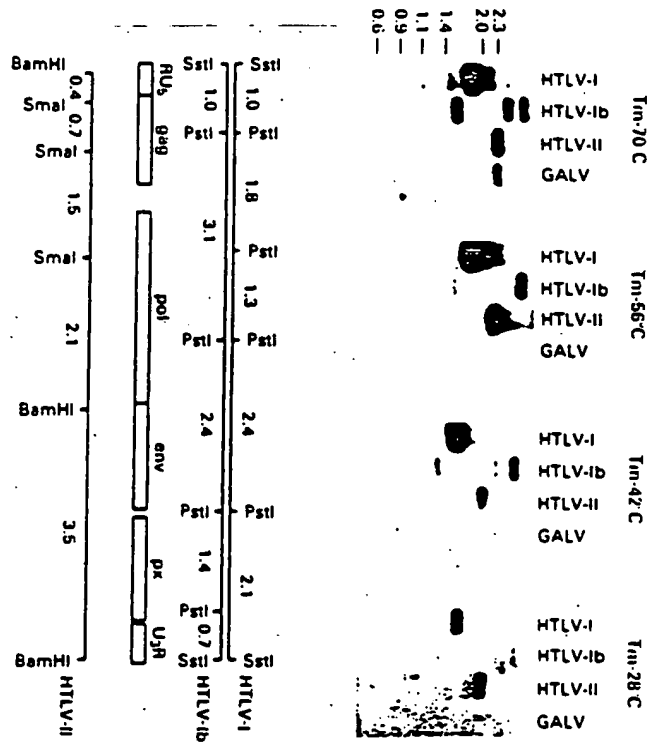


Fig 4 JWS 8/17/84
8/17/84
8/17/84
RG/ 8/17/84
MPJ 8/17/84

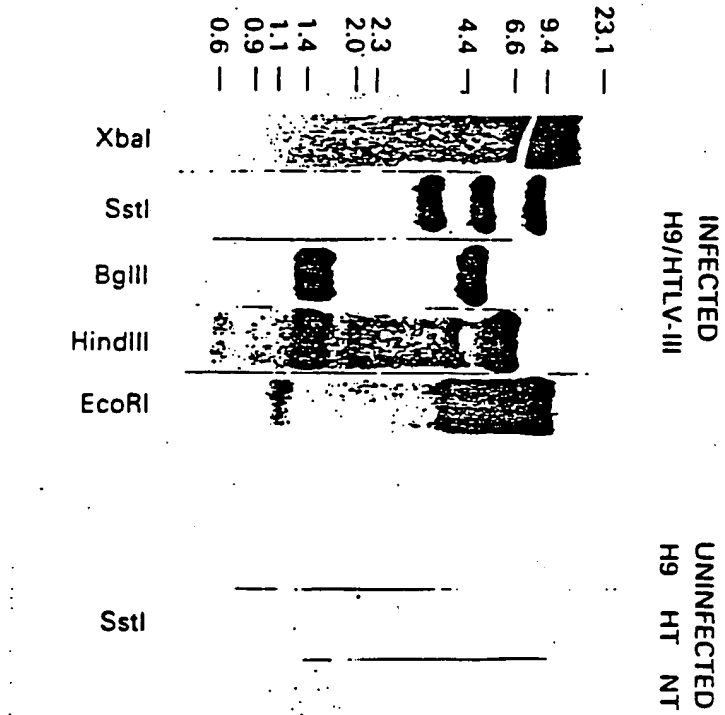


FIG 3 DWS 8/17/84

R6 8/17/84 8/18/84 MP/8/17/84

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

MOLECULAR CLONES OF THE GENOME OF HTLV-III

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on _____ as

Application Serial No. _____

and was amended on _____ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

None

Form PTO-FB-110 (8-83) Patent and Trademark Office-U.S. DEPARTMENT OF COMMERCE

Prior Foreign Application(s) — Priority Claimed

(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
None				
			☐	☐
			☐	☐
			☐	☐

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)
None		

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

English Language Declaration

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (*list name and registration number*)

John S. Roberts, Jr. Reg. No. 15,786

Suite 504
2001 Jefferson Davis Hwy.
Arlington, VA 22202

703 521-0600

Send Correspondence to:

John S. Roberts, Jr.

Suite 504, 2001 Jeff. Davis Hwy., Arlington, VA 22202

Direct Telephone Calls to: (*name and telephone number*)

John S. Roberts, Jr. 703 521-0600

Full name of sole or first inventor	FLOSSIE WONG-STAAL
Inventor's signature	Flossie Wong Staal ✓ Date 8/17/84
Residence	Potomac, Maryland
Citizenship	U.S.
Post Office Address	8418 Harker Dr., Potomac, Maryland 20854
Full name of second joint inventor, if any	ROBERT C. GALLO
Second inventor's signature	Robert C Gallo ✓ Date 8/17/84
Residence	Bethesda, Maryland
Citizenship	U.S.
Post Office Address	8513 Thornden Terrace, Bethesda, Maryland 20817

(Supply similar information and signature for third and subsequent joint inventors.)

Full name of third joint inventor

BEATRICE H. HAHN

Inventor's signature — Date

Beatrice Hahn — 8/17/84

Residence

Chevy Chase, Maryland

Citizenship

West Germany

Post Office Address

3123 Brooklawn Terrace, Chevy Chase, Maryland 20815

Nikolaus Tappeiner 8/17/84

3 Books Hill Rd #401

Bethesda, Md 20814

Czechoslovak

DIV. ____ CONT ____ GRANTED ____

INVENTOR(S): _____

_____

_____

INVENTION: _____

_____

FILING DATE: _____

SERIAL NO: _____ GROUP NO: ____

Do not write in above space

ASSIGNMENT (JOINT)
(Executive Order)

Mikulas
HAHN & Popovic

WHEREAS, we, FLOSSIE WONG-STAAL, ROBERT C. GALLO and BEATRICE H. / employees of the U.S. Public Health Service, Department of Health and Human Services, and citizens of the United States, have invented _____

MOLECULAR CLONES OF THE GENOME OF HTLV-III

for which we are about to make application, executed Aug. 17 1984 to the Commissioner of Patents for grant of Letters Patent of the United States; and

WHEREAS, we are the applicants named in the above identified application for Letters Patent; and

WHEREAS, the conditions under which said invention was made are such as to entitle the Government under Paragraph 1(a) of Executive Order 10096, to the entire right, title and interest therein, including foreign rights; and

WHEREAS, as to foreign rights, it is the policy of the Government to obtain an option to exercise such rights;

NOW, THEREFORE, to all whom it may concern; be it known that for and in consideration of the premises and other valuable considerations, we the undersigned, have sold, assigned and transferred and by these presents do sell, assign and transfer unto the Government of the United States of America as represented by the Secretary of the Department of Health and Human Services, the entire right, title and interest throughout the United States of America, its territories and dependencies, in and to the aforesaid invention described in the aforesaid application for Letters Patent of the United States, and all Letters Patent issuing thereon and any continuations, divisions and reissues or extensions thereof; hereby authorize and request the Commissioner of Patents to issue said Letters Patent to the Government of the United States of America, as represented by the Secretary of the Department of Health and Human Services, and his successors, as assignee of the entire right, title and interest in and to the same throughout the United States of America, its territories and dependencies, for the sole use for the full term or terms for which said Letters Patent and any continuations, divisions and reissues or extensions thereof are, or may be, granted as fully and entirely as the same would have been held by us, had this assignment not been made, and we do hereby grant unto the Government of the United States as represented by the Secretary of the Department of Health and Human Services, the option to take all of the right, title and interest in said invention or all applications for Letters Patent thereon in all countries foreign to the United States in which the Government of the United States may file, or cause to be filed, applications for Letters Patent, without payment to me of any further consideration; provided, however, that this grant of an option to take foreign rights in my invention, or applications for Letters Patent thereon, shall have force and effect only as to such applications filed in foreign countries within six months of the filing date of any applications for United States Letters Patent covering my invention, and that all foreign rights not exercised under the option are left to us subject to

INVENTOR:

Flossie Wong-Staal
FLOSSIE WONG-STAAL

SUBSCRIBED AND SWORN to before me this 17th day of August 19 84; at Bethesda in the County of Mo and State of Md

Rosemarie Magnuson
NOTARY

(SEAL)

My Commission Expires on July 1, 1986

INVENTOR:

Robert C. Gallo
ROBERT C. GALLO

SUBSCRIBED AND SWORN to before me this 17th day of August 19 84; at Bethesda in the County of _____ and State of Md

Rosemarie Magnuson
NOTARY

(SEAL)

My Commission Expires on July 1, 1986

INVENTOR:

Beatrice Hahn
BEATRICE H. HAHN

SUBSCRIBED AND SWORN to before me this 17th day of August 19 84; at Bethesda in the County of Mo and State of Md

Rosemarie Magnuson
NOTARY

(SEAL)

My Commission Expires on July 1, 1986

My Commission Expires on July 1, 1986

Subscribed and sworn to before me, in my presence, this 17th day of August 19 84 a Notary Public in and for the State of Maryland

Notary Public

My commission expires _____ 19 __

Rosemarie Magnuson

My Commission Expires on July 1, 1986